Registration & Modeling of Shapes with Uncertainties: Contributions and Applications to Knowledge Based Segmentation

Maxime Taron

THÈSE

présentée pour l'obtention du titre de

DOCTEUR DE L'ÉCOLE NATIONALE
DES PONTS ET CHAUSSÉES

Spécialité : Informatique

par

Maxime TARON

Recalage & Modélisation de Formes avec Incertitudes
Contributions et Applications à la Segmentation
avec a priori Statistique

Registration & Modeling of Shapes with Uncertainties:
Contributions and Applications to Knowledge Based Segmentation

Soutenue le 15 novembre 2007 devant le jury composé de :

Président du jury	Jérome Garot	CHU Henri Mondor - Paris XII
Rapporteurs	Michael J. Black	Brown University
	Isabelle Bloch	ENST
	Alison Noble	University of Oxford
Examinateurs	Hervé Delingette	INRIA
	Renaud Keriven	Ecole des Ponts
Directeurs de thèse	Nikos Paragios	Ecole Centrale de Paris
	Marie-Pierre Jolly	Siemens Corporate Research

Table of Contents

List of Figures

Acknowledgements

This Thesis work was performed mostly in Ecole Nationale des Ponts et Chaussées, and then Ecole Centrale de Paris. During these years, I was hosted by 3 research laboratories: CERTIS (" Centre d'Etudes et de Recherches sur les Techniques de l'Informations et des Systèmes " in Ecole de Ponts) , MAS (Applied Maths and Systems lab in Ecole Centrale) and the Imaging and Visualization department of Siemens Corporate Research who sponsored most of this research work. Also a PhD thesis is personal work, none of this would have been possible without the help of some people who counted most in the past 4 years. My greatest thanks go to Pr. Nikos Paragios, my PhD director, who guided me through the research work and helped me in keeping consistent motivation and faith which is the first requirement to go forward in research. Also I would like to thank my co-director in Siemens Corporate Research, Marie-Pierre Jolly who welcomed me in Princeton every summer. I greatly appreciated her constant support and advice on any matters.

Next, I would also like to express my deep thanks to Alison Noble, Isabelle Bloch and Mickael Black for having accepted to review my manuscript despite the important amount of work this required. Many thanks also to Jérôme Garot, Hervé Delingette and Renaud Keriven for their participation in my thesis committee.

Beside contributions and working matters, I also received a great amount of support from all my colleagues and friends. I would like to especially mention two of my office mates, Cedric Allene and Noura Azzabou. Also many other colleagues from Ecole des Ponts have counted in my first years of Phd, starting with Renaud Keriven, Geoffray Adde and Olivier Juan who have been a precious relief on technical issues. Jean Yves Audibert gave me crucial hints on the statistical part while Charlotte Ghys brought a very different application to the method presented in this work with the treatment of facial expressions. Also, I would like to acknowledge especially Patrick Etyngier for being my best reference frame in the last month of my Thesis, Radhouène Neji for reminding us of the irony of our lives; Stephane Soppera and Emeric Sarron for their friendship and gentle pressure while approaching the end of this thesis.

Introduction (Version Française)

La plupart des organismes vivants sont munis de centre de perception visuelle de plus ou moins grande complexité leur permettant d'interagir avec leur environnement immédiat. Le but principal de la vision par ordinateur consisterait donc à extraire et comprendre le contenu d'images afin de reproduire artificiellement cette capacité d'interaction à l'aide de robots ou de bras motorisés. Cependant l'interaction physique n'est pas essentielle, un but annexe tend à utiliser ces techniques afin d'améliorer la qualité des liens homme/ordinateurs avec des applications liées au traitement des images, des flux vidéo (surveillance, suivi d'objets et structures...) ou encore l'aide au diagnostic dans le milieu médical. Cette problématique d'interprétation des images fait donc appel à plusieurs domaines de recherche tels que la reconnaissance de structures, intelligence artificielle, vision par ordinateur, traitement des images, apprentissage, etc. Plus récemment l'analyse des images médicales est apparue comme un axe de recherche important faisant appel à plusieurs des branches précédemment évoquées. Depuis quarante ans les travaux de recherche sur ces sujets sont nombreux et touchent des domaines scientifiques aussi vaste que les mathématiques appliquées, l'informatique, la mécanique, la physiologie, biologie et l'anatomie.

Supposant que le système visuel de l'être humain est le plus performant (en termes d'acquisition et de traitement), de nombreux travaux ont été effectués afin d'en comprendre son fonctionnement et de pouvoir le reproduire en partie. Pourtant l'état de l'art sur la compréhension du système humain découlant des théories de David Marr, reste insuffisant pour pouvoir être appliqué directement et l'essentiel des recherches applicatives de la vision par ordinateur vont principalement se focaliser sur différentes techniques issues des mathématiques appliquées. De telles approches nécessitent la sélection d'un modèle (représentation paramétrique du problème) établissant un lien avec les observations (images, données médicales, sorties de systèmes stéréoscopiques ou multi vues, etc.) ainsi que l'utilisation de techniques efficaces pour estimer les paramètres du modèle. Il existe donc une infinité d'approches possibles pour résoudre un problème de perception visuelle. Par ailleurs ces problèmes sont très souvent mal posés car les dépendances existant entre observations et modèle sont non linéaires, faiblement contraintes et donc difficiles à optimiser.

Par ailleurs, le choix de l'approche utilisée pour résoudre un problème lié à la perception visuelle est largement justifié par le domaine d'application. Par exemple, le positionnement automatique et la localisation d'obstacles seront considérés en robotique, l'estimation des déplacements et le suivi d'objets sont utilisés pour traiter les flux vidéos et les techniques de reconstruction ou d'estimation de la position d'objets tridimensionnels sont étudiés lorsque des systèmes stéréoscopique ou monoculaire sont utilisés. Enfin, des modèles statistiques seront souvent utilisés en traitement des images médicales.

Durant les deux dernières décennies, nous avons pu observer d'importants progrès dans les techniques permettant d'examiner les tissus biologiques de façon non invasive. Les dernières générations de scanners fournissent des informations anatomiques et physiologiques ainsi qu'un faisceau de données liées à certaines pathologies pouvant être exploitées pour un diagnostic précoce et suivies de mesures thérapeutiques adaptées. La généralisation des techniques d'imageries médicales, l'amélioration de la qualité des images et l'introduction de nouvelles modalités ont contribué à produire une large quantité de données devant être analysées. Ceci a rendu à la fois indispensable et réalisable les techniques automatiques de prétraitement et autres outils d'évaluation assistés par ordinateur sur les images.

L'imagerie médicale fait référence à un ensemble de modalités décrivant l'état des tissus et des os, ainsi que diverse informations physiologiques. Les différentes techniques d'acquisitions parmi lesquelles comptent les Rayons X, Scanner, Imagerie à résonance magnétique, échographie, imagerie à tenseur de diffusion, tomographie à émissions de positrons etc. présentent des propriétés complémentaires. L'exploitation des informations issues de ces modalités constitue un défi moderne pour comprendre l'anatomie, les structures biologiques et les effets des pathologies sur ce qui est visualisable de leur fonctionnement (au travers des déplacements, diffusions de molécules marquées, flux sanguin, activité neuronale etc.) Finalement, l'aide au diagnostic médical peut être résumé par le traitement de ces signaux multidimensionnels structurés permettant de comprendre l'état des organes étudiés. Une telle problématique est clairement pluridisciplinaire et fait appel à des domaines scientifiques allant de la biologie aux mathématiques.

Contexte et Motivations

La modélisation de structures anatomiques complexes se décompose généralement en trois étapes: (i) déterminer un ensemble de mesures anatomiques et pathologiques obtenues grâce au traitement et l'exploitation des différentes modalités d'imagerie médicales. (ii) proposer un modèle paramétrique en accord avec les indices anatomiques capable de décrire l'ensemble des variations

de l'organe considéré; (iii) les paramètres de ce modèle sont estimés afin de reproduire le comportement d'une nouvelle observation sur laquelle les mêmes indices anatomiques et/ou pathologiques sont présents. On peut alors comparer cette nouvelle observation avec celle d'un cas de référence, sain, afin d'obtenir des indications aidant au diagnostic. Ce procédé est généralement accompagné par l'étude d'un modèle statistique décrivant les variations de l'organe, construit à partir d'individus sains et malades afin de le comparer plus efficacement à une nouvelle observation.

Les techniques de représentation de formes et l'étude de leurs variations pour des objets appartenant à la même classe ou a des classes différentes sont des problèmes universels aux applications multiples. Les principales applications sont la segmentation avec a priori statistique de formes (l'extraction de structures particulières sur des images), le 'tracking' (suivi d'objets, l'obtention des positions et déformations successives d'une structure dans une séquence d'images), la reconnaissance d'objets (classification permettant de décider si un certain type d'objet est présent dans une image), etc. L'extraction des informations caractérisant l'état d'un organe reste un objectif ambitieux. Il s'agit d'abord d'interpréter les données en provenance des signaux source, et d'en fusionner les informations. Ces informations peuvent prendre la forme de signaux discrets tels que des images, volumes 3D ou 4D, sur lesquels les signaux présentent un certain degré de corrélation à différentes échelles. Retrouver le contenu sémantique d'une image consiste alors en la description de ces mesures à l'aide d'un modèle mathématique simple dont on estimera les paramètres. Si un tel procédé peut sembler simpliste, il faut en mesurer la complexité au regard des organes à modéliser, pour lesquels la sélection d'un modèle mathématique résulte d'un compromis entre complexité et faisabilité. D'une part, on cherchera à créer un modèle capable d'exprimer les différents états de l'organe étudié; et d'autre part, on déterminera les valeurs des paramètres de ce modèle uniquement à partir de nouvelles mesures pouvant s'avérer incomplètes. Considérons un exemple simple tiré du cycle cardiaque. Le coeur est un muscle possédant deux ventricules. Le ventricule gauche, le plus gros et le plus important agit comme une pompe et envoie le sang enrichi en oxygène vers tous les tissus et organes. Le cycle cardiaque présente deux états essentiels: La fin de la diastole, correspondant à l'instant où les ventricules sont le plus dilatés, et la fin de la systole, l'instant où le coeur est le plus contracté. La différence entre les deux volumes ventriculaires à ces deux instants est un indice physiologique important sur le fonctionnement cardiaque. Dans cette optique, l'extraction de données fait référence à l'obtention du volume de la cavité du ventricule gauche donc à sa segmentation.

Dès lors que ces informations ont été extraites des images, l'étape suivante résidera dans la sélection du modèle en vue de la séparation entre les cas normaux et les cas affectés par diverses pathologies. Ceci est généralement effectué via un modèle statistique décrivant les différentes populations. Cette idée relativement simple suppose qu'un ensemble de mesures ont été extraites

sur différents individus et étudie les différences existant entre ces mesures. L'objectif de la phase de modélisation est de retrouver des comportements analogues entre des individus, qui pourront ensuite être utilisés comme autant de références dans l'étude de nouveaux cas. Ceci reste une vue simpliste de la tâche à effectuer ; dans la pratique des modèles plus complexes provenant de la théorie de l'estimation où des statistiques seront utilisés. Par ailleurs, l'ensemble des observations est contraint par le modèle anatomique choisi. Si on considère encore une fois l'exemple du cœur et de la fraction d'éjection, on utilisera le fait que cette quantité pour un cœur sain est d'environ 60

Dès lors que ces modèles ont été déterminés et que leurs statistiques ont été estimées/apprises à partir des observations, le diagnostic est le résultat de la comparaison entre les mesures effectuées sur un nouveau cas et les comportements théoriques appris à partir de l'ensemble des cas de référence. Cette étape nécessite encore une fois l'extraction d'informations à partir d'image médicale et a pour objectif d'informer des risques potentiels encourus par le patient. Il faut néanmoins être conscient que le développement de telles technologies n'a pas pour but de remplacer ou marginaliser le rôle du médecin : le traitement des images médicales, pour l'aide au diagnostic, doit se limiter à fournir au médecin un panel d'outils appropriés lui permettant de faire un diagnostic plus précis, prendre des décisions plus rapidement tout en minimisant le risque pour le patient.

Parmi l'ensemble des étapes présentées précédemment pour l'aide au diagnostic, la plus complexe est la construction du modèle mathématique de l'organe étudié. Cette étape a d'ailleurs un impact important sur les autres composantes du processus. Tout d'abord ce modèle est utilisé pour aider à l'extraction de l'organe à partir des images biomédicales ; ensuite, ce modèle constitue la base des comparaisons entre les sujets sains et malades. Les méthodes présentées dans l'état de l'art, séparent généralement les étapes de modélisation d'un organe et l'étape d'extraction de cet organe dans les images. En d'autres termes, ces méthodes considèrent que l'on dispose d'un ensemble d'apprentissages et d'un modèle statistique dont les paramètres sont alors estimés. Par ailleurs, le recalage constitue aussi un élément déterminant (le recalage est l'étape durant laquelle tous les éléments de l'ensemble d'apprentissages sont transformés pour pouvoir être représentés dans le même référentiel). Le recalage pouvant aussi être l'objet d'erreurs importantes, il donne lieu à des modèles de formes erronés pour lesquels les erreurs d'alignement se sont propagées. La sélection d'un modèle pour les déformations est aussi déterminante, il s'agit de créer un modèle suffisamment simple pour pouvoir en estimer les paramètres de façon robuste, et suffisamment complexe pour représenter l'ensemble des déformations possibles, qu'ils correspondent à des cas normaux (sains) ou anormaux (ayant subi diverses affections). Enfin ce modèle statistique sera considéré pour aider au processus de segmentation, dans les cas où les informations contenues dans les images médicales ne sont pas suffisantes pour réaliser une segmentation précise. L'utilisation

d'un tel modèle dans le processus de segmentation est la pierre angulaire de l'aide au diagnostic médical.

Contributions

Cette thèse a permis de traiter les problèmes essentiels de la vision par ordinateur que constitue le recalage des surfaces (global et local), la modélisation de formes, et la segmentation avec a priori de forme. Dans cette optique nous avons considéré une technique de représentation de la forme basée sur le calcul de cartes de distances. Cette technique bénéficie de nombreux avantages tant géométriques que mathématiques et sera utilisé tout au long de la chaîne de raisonnement.

Le recalage de formes ou de surfaces, est un problème central de la vision par ordinateur, objet de nombreuses études et souvent utilisé dans le traitement des images médicales. Ce problème se résume à déterminer une déformation particulière permettant d'établir l'ensemble des correspondances entre deux formes ou deux ensembles de caractéristiques extraites sur des images. Ces deux objets étant respectivement nommés la 'source' (pouvant être déformée) et la 'cible' sur laquelle la source est déformée, Dans le cadre de notre étude, les ensembles considérés seront des formes représentés à l'aide de cartes de distance. La contribution la plus importante de nos travaux réside dans l'introduction et l'utilisation des incertitudes dans le processus de recalage. Ainsi, l'action d'aligner deux surfaces n'est plus considéré comme un problème à solution unique, mais comme ayant une infinité de solutions dont la distribution pourrait être représentée par une densité de probabilité multimodale, dans un espace de grande dimension. Une telle approche considérée au niveau local, dans un voisinage de la solution de recalage obtenue, fournit une information quantitative sur la qualité du résultat. En d'autres termes, si on considère une surface source alignée sur une surface cible de façon globale, le recalage local permet d'établir un ensemble de correspondances entre les surfaces, et les incertitudes d'introduire un ensemble de matrices de covariance indiquant le degré de confiance pouvant être donné à ces correspondances. Le problème de recalage a été formulé de façon très générique, afin de le réaliser nous avons considéré dans un premier temps une classe de déformations libres (FFD, " Free Form Deformations " ou déformations de formes libres) permettant de représenter une déformation de l'espace ambiant indépendamment de la forme ou surface considérée. Dans un second temps les déformations de plaques minces ont été considérées (TPS, " Thin Plate Spline ", ou déformations des plaques minces), elles permettent une déformation des surfaces et induit une déformation de tout l'espace, la particularité réside dans la position des poignées contrôlant la déformation (ou points de contrôle), placées à la surface de l'organe ce qui en fait une transformation dépendant spécifiquement de l'organe, mais ayant une plus faible dimension que les déformations libres pour une précision équivalente. Ces transfor-

mations permettent de représenter indifféremment des transformations locales ou globales, nous proposons donc d'en séparer la composante rigide indispensable dans la phase de modélisation.

Dès lors que le problème du recalage a été réglé, il s'agit de modéliser les variations présentes dans un ensemble d'apprentissages. Concrètement, il faut déterminer une densité de probabilité pouvant représenter ces variations. Notre seconde principale contribution réside dans l'approche utilisée pour propager les incertitudes évaluées pendant le recalage dans la phase de modélisation. Nous avons développé deux approches indépendantes. La première génère un modèle de faible dimension utilisant les techniques développées pour l'analyse en composante indépendante (ACI) et adapté à la modélisation des déformations. Cette approche est utilisée et hérite de l'information provenant des incertitudes. Pour cela, l'ensemble d'apprentissages est utilisé comme une base pour générer de nombreux nouveaux échantillons, de sorte que les exemples avec une forte variance seront plus largement dispersés, donnant un plus faible degré de confiance à ces variations. Le calcul de l'analyse en composante indépendante avec les incertitudes est un modèle efficace si les variations des échantillons présentent des variations indépendantes pouvant s'exprimer comme des mixtures de Gaussiennes. Pourtant une telle contrainte n'est pas systématiquement réalisée sur des cas pratiques ; afin de pallier ce défaut, nous avons proposé une approche non paramétrique utilisant des noyaux avec une covariance variable afin de modéliser les variations des échantillons. L'idée essentielle considérera que les échantillons les plus représentatifs seront associés à des variables aléatoires Gaussienne dans l'espace des déformations, centrées sur la déformation et donc la matrice de covariance dépend de l'incertitude locale de l'échantillon considéré.

Ayant ainsi construit deux modèles statistiques pour ces déformations, ils seront utilisés pour effectuer une segmentation des organes sur des images médicales avec un a priori de forme. Pour cela, nous combinons une représentation implicite des formes (avec des ensembles de niveaux ou level sets), avec des déformations paramétriques basées sur un ensemble de points de contrôle. Les points de contrôle pourront être localisés sur une grille régulière (déformation de formes libres, FFD) ou sur une surface (déformation de plaques minces, TPS). Le modèle de moindre complexité utilise conjointement les TPS et l'analyse en composante indépendante pour la segmentation du ventricule gauche en imagerie scanner tridimensionnelle. Pour modéliser des organes qui présentent des variations plus importantes ne montrant pas de structure et de régularité, nous avons utilisé l'approche non paramétrique à noyaux de taille variable associée aux déformations de formes libres (FFD). Le corps calleux a été utilisé comme exemple test pour cette approche, la segmentation étant considérée sur des vues sagittales en imagerie à résonance magnétique. Dans les deux cas considérés, le résultat final de la segmentation a une interprétation qualitative et quantitative. Ainsi, le recalage tout comme la segmentation sont vues comme des problèmes d'estimation, les incertitudes apportant des informations relatives aux données disponibles pour

effectuer cette tâche (support visuel des images en segmentation et bruit inhérent aux formes et surfaces pour le recalage). Enfin, ces mesures d'incertitudes contribuent à la segmentation, dans une approche indépendante du choix des paramètres et permettant de retrouver d'importantes variations de formes.

La thèse suit ce schéma et est découpée en trois grands chapitres. D'abord le recalage, introduisant la technique de représentation des formes, les différentes transformations, les techniques d'optimisation et d'évaluation des résultats. Ce chapitre introduit donc les mesures d'incertitudes en proposant deux approches complémentaires. Le second chapitre touche à la modélisation des formes, introduisant les modèles paramétriques et non paramétriques et l'usage des incertitudes dans ces modèles. Le troisième chapitre décrit l'utilisation de ces méthodes pour la segmentation et introduit les incertitudes dans le schéma d'évolution du modèle déformable ainsi introduit. Le dernier chapitre clos la thèse en introduisant une série de directions à venir que peut prendre ces travaux. Enfin les appendices successivement le calcul des dérivées utilisées dans la phase de segmentation puis une application important du modèle statistique pour une application non médicale.

Chapter 1

Introduction

Visual perception is a central component for most biological organisms. Despite an important degree of variation between natural organisms, even the most primitive element has some abilities to sense the environment through visual sensors. Understanding and extracting content from images and reproducing this ability in artificial environments like intelligent robots, robotic arms, computer aided interpretation of images and video, human computer interaction, computer aided diagnosis, etc. is the primary goal of computer vision. Statistical pattern recognition, artificial intelligence robotics, computer vision as well as image processing and more recently medical image analysis are research domains which related visual perception to image understanding. These domains have gained significant attention from the scientific community in the past four decades and now are established and well represented research areas in applied mathematics, computer and engineering sciences, physiology, biology and neurology. One can observe an interdisciplinary effort from scientists with diverse scientific backgrounds towards visual perception.

Under the certitude that the human vision system (acquisition as well as processing) is the most efficient visual perception system, recent efforts in the above mentioned vision related areas aim to reproduce human visual system. However, the current state of understanding of the human brain is far from being reproducible. Therefore, the available knowledge is not sufficient to replicate it, through the aid of sensors and computers complex biological vision systems. The main stream of research over the past decades in the area of computer vision lies mostly in the use of applied mathematics and engineering. In such a context, research scientists have abandoned David Marr's theories and introduced parametric mathematical inference problem for specific tasks of visual perception, where the set of optimal parameters corresponds to the answer of visual perception. This process requires the selection of a model (parametric representation of the problem), the establishment of a connection between the model and the observations (images, medical data, video,

stereo data, etc.) and the use of efficient strategies to recover the model parameters. One should point out that such strategies have an infinite number of possible solutions since one can formalize the same perception tasks with numerous mathematical models, use various means of introducing the model to data dependencies as well as different techniques exist for the optimization of these dependencies. Furthermore, most of these approaches are ill-posed due to the non-linearity of the dependencies between parameters and observations, as well as due to the lack of sufficient constraints to guide the estimation problem.

Therefore, the definition of a universal solution to the visual perception problem is rather challenging and strongly depends on the task to be accomplished. The task definition is also strongly related to the application area. For example, self-localization and obstacle avoidance are problems often considered in robotics, motion estimation and tracking are studied in video processing, stereo reconstruction and 3D inference are investigated in computer vision, statistical modeling of samples and segmentation at looked at in medical image analysis.

In the recent years, we have observed a revolution on how human and biological tissues can be imaged in non-invasive ways. The latest generation of medical hardware provides anatomical, physiological and pathological data which can be used to perform early diagnosis, follow up and evaluation of therapeutic strategies. Constant improvements on the image quality as well as the introduction of new image modalities have generated an enormous amount of data to be analized. In such a context, the use of computer aided-techniques has emerged as an efficient pre-screening procedure with applications to diagnosis and post-treatment evaluation.

Medical images refer to a set of modalities describing the status of human tissues, bones and physiological information. This may refer to simple or very complex measurements and can be acquired in a number of ways. X-rays, Magnetic Resonance, Computer Tomography, Positron Emission Tomography, Diffusion Tensor Imaging are examples of acquisitions with varying and complementary properties. The exploitation of such an information space is a great challenge of our days and consists of understanding the anatomical structure of biological systems and in particular the effect of pathologies on their complex mechanisms of operation. The task of computer aided diagnosis can be reformulated as follows: processing these n-dimensional signals in order to understand the current state of an organ of interest. Such in-depth modeling and understanding of complex biological systems is an interdisciplinary effort which involves researchers with different scientific backgrounds including physiology, biology, neurobiology, mathematics and engineering.

1.1 Context and Motivations

Modeling complex anatomical structures often consists of three steps; (i) A set of measures, of anatomical and pathological indices are recovered through the processing, understanding and exploitation of medical image modalities, (ii) A (parametric) mathematical model that is consistent with the anatomy is proposed, it is capable of describing the operation (i.e. the statistical variation) of the organ or structure under consideration, and (iii) estimating the parameters of the model so that it can reproduce the behavior observed through the use of anatomical and pathological indices. Then one may compare this behavior with a reference, healthy behavior, to make a soft diagnosis. Such a task is achieved through a statistical comparison between the model built to describe healthy individuals and the measurements obtained for the subject under consideration.

Shape representation and modeling of its variations inter and intra-class is considered to be a universal problem with applications to knowledge-based segmentation (extraction from images of a particular structure), tracking (recovering successive positions and deformations of a structure in a number of consecutive images), recognition (classification decision for the presence or not of a class of objects in an observed image), etc. Extraction of information regarding the state of an organ from biomedical images is a challenging task. One has to deal first with content interpretation from sparse local signals and then with the fusion problem. Images, volumes, 4D volumes etc. correspond to sampled continuous functions where measurements are only correlated at a local scale, therefore recovering content often consists of describing the measurements using a simple mathematical model and then estimating the parameters of this model using the measurements. While such a process seems very trivial, given the complexity of biological systems the selection of a model which is a compromise between complexity and tractability is not straightforward. On one hand, one would like a model that is capable of explaining the state of the organ; on the other hand, one should be able to determine the parameters of this model from sparse signals. Let us consider a ”simple” example: the heart cycle and in particular the cycle of the left ventricle which pumps oxygenated blood to all human tissues (including the most distant one). The cardiac cycle contains two important phases alternating contraction and dilatation. End-systole and end-diastole are the instants where the ventricle has its lowest and its highest volume. The relative volumetric difference between these instants (called ejection fraction) is a reliable indicator about the heart function. In such a context the task of content extraction refers to the recovery of the volume of the left ventricle.

Once information has been determined from images, the next step consists of determining appropriate means to model this information and being able to separate the diseased from the normal

cases. This is often done through statistical modeling of populations. The idea is fairly simple: assuming a set of measurements from different individuals, study how different these measurements are. The aim of modeling is to recover a common pattern of behavior for the observations between subjects which then can be used as a gold standard to compare with new cases. This is a rather simplistic representation of the task. In practice complex mathematical models from estimation theory as well as statistical inference are considered to construct these models. The observations are constrained from known anatomical models. In order to demonstrate this task, let us again consider the heart operation. It is known and will be confirmed that the ejection fraction should be about 60%, which means that the blood leaving the heart in every cardiac cycle is about 60% of the end-diastolic left ventricle volume. Eventually this is a rather simplistic example since the measurements refer to a single dimension, while in the most general case one can imagine measurements of high dimension with many indices to be statistically modeled.

Once models have been determined and have been estimated from observations, diagnosis refers to the task of comparing the measurement of the new subject with the expected theoretical behavior of the model. This task involves again content extraction from biomedical images and aims to produce an indicator of whether the subject under consideration has a potential risk. One should have in mind that this technology does not aim to replace or marginalize the role of physicians. The task of computer aided diagnosis consists in providing more appropriate means of content interpretation in medical images, which will allow faster and more accurate diagnosis, while producing new means for treatment and therapy evaluation.

Efficient mathematical organ representation is among the most challenging problems of the above mentioned processing chain because it has a large impact on all other components of the process. In particular it can be used as an aid to improve organ extraction from images, and also is the base of comparison between healthy and unhealthy subjects. State of the art methods often dis-associate the problem of organ modeling with organ extraction. In other words, given a set of training examples and a choice of the statistical model, they infer the model parameters. Registration itself (bringing all examples of the training set in the same parameter space) is challenging and often erroneous, resulting in models where alignment errors have been propagated. Furthermore, the selection of a mathematical model to account for the variation of samples is also critical. First, one would like a simple mathematical model where robust parameter estimation is feasible. Furthermore, a need exists for a model capable of accommodating the variation of all samples while being able to cope with abnormal samples. Last, but not least these models should be considered efficiently to aid the segmentation process if the data support is not available and provide efficient diagnosis.

1.2 Contributions

In this thesis we aim to address the challenges that are surface registration (global and local), modeling shape variations (parametric and non parametric models) and knowledge-based segmentation. To this end, we consider a state-of-the art shape/surface representation (distance transforms) with numerous desirable geometrical and mathematical properties which is then used along the proposed chain.

Shape/surface registration is a well studied problem in computer vision, computational geometry and medical image analysis. The definition of the problem consists of deriving a transformation, which given a source and a target feature space, establishes correspondences between them. In the context of our research, feature spaces correspond to 2D curves and 3D shapes, represented using distance transforms. The main contribution of our work consists of the introduction of uncertainties in the registration process. Therefore, the transformation aligning the two surfaces becomes a multi-modal high-dimensional density function with both quantitative and qualitative expression of the process. In other words, given a source and a target representation as well as an initial global alignment between them, we are able to determine the set of correspondences and also associate the result with covariance matrices which indicate the amount of confidence for the obtained result. In order to address generic surface registration this concept is customized for the case of free form deformations (FFD) which is a shape/surface free (domain-defined) representation of dense displacements. Next, we propose a similar estimation of a multivariate deformation density for the case of thin plate splines representations (TPS). This is an organ specific, low-dimensional transformation with control-points being defined on the surface. In order to account for the integrated nature of this transformation (the global and the local are not separable), we propose a rigid-invariant form.

Once the registration problem has been addressed, the next task consists of modeling the variations of the training set, or determining a continuous probabilistic representation of the observed density. The main contribution of our work consists of presenting two alternatives on propagating registration uncertainties to the statistical model describing shape variations. The first consists of generating a reduced model that uses independent component analysis (ICA) to model the computed class of transformations. The traditional independent component analysis approach is improved and inherits the registration uncertainties through the augmentation of the sample set according to the observed uncertainties. Therefore, samples with high variance are dispersed and produce statistical shape models which are less confident in these areas. The uncertainty-driven ICA is an efficient computational statistical model if samples can be expressed as a mixture of

Gaussians. However, there are cases where imposing such a parametric constraint is not natural and the observed shape variation cannot be quantified using these models. In order to overcome this limitation, we propose a variable-bandwidth non-parametric approximation of the samples. The central idea behind such an approach is to consider the most representative examples in the training set and associate them with continuous Gaussian densities centered at the deformation with a covariance matrix that depends on the local uncertainty of the considered sample.

With these statistical models in hand, the next task to be addressed is knowledge-based segmentation and shape-based classification. Both statistical models are used to impose prior knowledge on the segmentation. We efficiently combine implicit parameter-free representations (level sets), with parametric control-point based deformations either defined on a regular grid (FFD) or considered on a surface grid (TPS). The model with the least complexity associates thin plate splines deformations and independent component analysis and is used to segment the left ventricle in cardiac computer tomography and magnetic resonance images. For organs with larger shape variations we propose the variable bandwidth non-parametric (VBD) approach associated with free form deformations (FFD). The Corpus Callosum is used as a test a case for shape-driven segmentation. In both cases, the final segmentation result has a qualitative and a quantitative interpretation. Segmentation like registration is viewed as a statistical estimation problem where uncertainties can be determined which encode both the amount of support from the prior model as well as the local image support. Last, but not least in order to improve classification, both the deformation and the uncertainty are considered in such a statistical model which is relatively parameter-free and can capture important variations of the training examples.

The remainder of this document is organized in three main chapters, a registration one, a statistical modeling one, and a segmentation one. Conclusions are part of that last section while non-related medical applications are presented in the appendix along with derivations of cost functions which were presented in the main document.

1.3 Thesis Plan

Chapter 2 is dedicated to shape registration which involves three aspects, Shape representations, nature and type of transformations and optimization strategies. We first review the most widely used shape representations and motivate our selection to use distance functions. Once the representation has been established, we discuss/review the nature of deformations in particular focusing on dense parametric registration models. Two models are presented in details: free form deformations and thin plate spline deformations. Then, we briefly review the state-of-the art point-based regis-

tration methods (Iterated Closed Point, Shape Context, and Robust-Point-Matching) and introduce our approach on the space of implicit functions focusing on the explicit estimation of uncertainties which is the main contribution of this chapter for both deformations models being considered.

In the next chapter we first present the state of the art in modeling shape variations which often requires the selection of the parametric model and the estimation of its parameters from the data. In particular we briefly review simple models like principal component analysis, mixture of Gaussians and kernel principal component analysis among others. Then we present in details the independent component analysis approach and then non-parametric density approximations with fixed and variable bandwidths. These models are then modified and enriched with the uncertainty-driven registration results presented earlier leading to more efficient representations of densities which can better capture the samples variation and can perform more efficient samples discrimination/classification.

The last main chapter of the thesis is dedicated to knowledge-based image segmentation. First, we review model-free and model-based image segmentation as well as their applications to computer vision. In particular we focus on level-set based methods, as well as active shape models which are the most closely related with our approach. Then we introduce two novel prior models, one based on uncertainty-driven independent component analysis and one on variable metric kernels. For both models we consider a state-of-the-art image term which can be adjusted according to the application setting and explicitly determine segmentation uncertainties. The proposed methods outperform complex approaches based on level sets as well as the ones with limited statistical capture (Gaussian assumptions) like active shape and appearance models. We demonstrate the performance of these methods using two applications, namely the segmentation of the left ventricle in cardiac images and the segmentation of the corpus callosum in brain images.

Conclusions and discussions are part of the last section of the document presenting first the main shortcoming and limitations of the approaches and then potential future perspectives. The document also contains an appendix where some of the contributions of the thesis are put in evidence for other application domains like statistical character recognition, facial animations and image morphing.

To conclude, this thesis evolves around shape, representation, registration, modeling, and image-based shape inference with constraints and well as primarily their applications in medical image analysis. It has produced (up to now), one major journal publication [185], one book chapter [136] two major conference publications [182, 181], two major workshop papers [184, 183] and three US patents (pending).

Chapter 2

Shape Registration

Abstract – *Shape registration consists of recovering a transformation that establishes certain correspondence between two structures (curves/surfaces/etc.). The problem is often ill posed because (i) the structures of interest can be represented in various ways, (ii) the set of allowable transformations is infinite, and (iii) the similarity between registered structures can be defined using various metrics. These 3 aspects of the registration problem are often treated as a whole, because the definition of shape takes multiple aspects in the state of the art and may carry deformations.*
This chapter will propose a definition of shape that meets the requirements of most of the descriptions and representations that will be reviewed. Then we will show the advantage of using an implicit shape representation with a distance transform as an implicit deformable template for registration. The chapter will then present the class of parametric deformations used in the registration process. Finally we will introduce the uncertainty on the registration results as a way of expressing the variability on the retrieved deformation field.

2.1 Introduction

The problem of shape registration can be simply defined as follows: given two instances of the same object, find a transformation that will align them according to some similarity measure. It has been a problem widely studied over the past decades with many related applications. The above-mentioned definition involves three terms, shape instance, shape transformation and shape similarity/disimilarity. Shape model or shape parameterization is often the term used to describe the mathematical model behind the physical object. Such a model aims to capture certain mathematical properties and often provides a continuous reconstruction. Transformation is the class of plausible geometric deformations which can be applied to the source shape towards improving the alignment of its transformed variant and the target shape. Last, but not least similarity is a mathematical term which defines measures of comparison between the transformed source shape and the target one.

The problem of efficient shape representations arises in numerous scientific domains like computational geometry, computer vision, graphics and animations, pattern recognition, computational and molecular biology, etc. The main aim is to determine an efficient mathematical model with nice computational properties which can be easily extracted and interpreted from a computer. These two desirable properties are often conflicting. Existing representations of shapes can be roughly classified in three categories which are representations of increasing complexity: (i) moments-based, (ii) discrete models, and (iii) continuous models [121]. Moment-based representations focus on the observed local structure of the shape [101]. In such a context one seeks for moments (preferably invariants) which can be either extracted from the object silhouette or from the object area. These moments encode the local characteristics of the shape and do not always retain the global structure. Therefore they are efficient for shape matching, recognition, etc. but have little success when considered for either registration or segmentation due to the fact that the inverse projection between these moments and the original shape is usually not trivial [215]. Discrete approaches exploit shape representations through a limited number of control points and some interpolation strategies, towards a complete geometric representation. One can then extract continuous geometric variables from this representation. The use of dense sampling leads to models of finer and finer precision that decreases the importance of the triangulation process, a widely considered technique to link control points. Subdivision surfaces are another example where the same concept is used. The main limitation of these representations lies in the fact that for the case of a finite number of control points, an important cross-dependency exists between the reconstructed shape and the sampling rule used to determine the positions of these points [206]. Furthermore, determining the geometric/physical properties of the shape either at a global or a local scale is far from trivial. Continuous models

are a geometric alternative to discrete models where one can recover a continuous/differentiable shape form through the use of interpolation functions. One can differentiate explicit from implicit representations in this category. In the former case, shapes are represented using either a number of control points and a continuous interpolation function or through a linear composition of a set of orthogonal features [37]. These methods are a reasonable compromise between complexity and geometric efficiency, but they suffer from their inability to describe forms with multiple non-connected components. Implicit representations are a useful set of alternatives, where a shape is represented using the zero-level set of a continuous function. These representations inherit some nice invariance properties, are geometrically efficient but computationally expensive [16].

The transformation that relates two instances is in general unrelated to the parameterization that was considered to represent the source and the target shape. In general, one can decompose prior art in global registration and dense alignment [167, 197]. Global registration assumes that all shape elements (point clouds, control points, triangulated surfaces) are transformed with the same parametric model, which in most of the cases is customized to them by being a function of their position. Translation or more complex models like rigid, similarity, affine or homographic are examples of global models [88]. The estimation of the transformation parameters is often a well-posed problem due to its over-constrained nature. Each shape element will provide a constraint while the same unknown transformation parameters are to be determined from all constraints. On the other hand these models fail to capture local deformations and they can be very imprecise if large local changes are observed between the two shape instances. Local deformation models are the most appropriate tools for local alignment. These models assume that either points move individually (extreme case), or they undergo a rich deformation according to a predefined mathematical model [191, 172]. In the former case the estimation of the individual deformations is rather ill-posed and therefore additional regularization constraints are to be considered for the recovery of a meaningful solution. Continuous deformation models are an alternative to point-based dense registration. With spline-based deformation, the transformation is expressed as a combination of a limited number of basis functions. In order to account for local consistency, these models also inherit some elasticity properties leading to continuous deformation models [58]. Examples of these models are free form deformations, thin plate splines, finite elements, etc. These models offer a compromise between global and dense registration with the estimation of their parameters being a tractable computational problem. However, one should note that their performance heavily depends on whether or not the selected continuous model is capable of expressing the observed deformation. In particular the use of elasticity constraints which is required towards recovering a geometrically meaningful solution could have an important impact on the quality of the obtained result [209].

The definition of a similarity/dissimilarity metric is strongly related to both the shape parameterization as well as the nature of the transformation. The Euclidean distance is the most natural metric when considering clouds of points independently of the nature of the transformation [208]. More advanced geometric features including normal, curvature or other higher order moments between shapes can be used when a continuous representation is assumed independently from the nature of the transformation [5]. These methods often decompose the estimation process in two stages, first correspondences between shapes is retrieved and then the transformation that aligns them is estimated [217]. The correspondences are often recovered on an expanded feature space (local curvature, etc.) while the alignment is done using geometric distances. The iterated closest point, the robust point matching and the dual-bootstrap iterated closest point are examples of such methods [13, 175]. Implicit representations of shapes offer natural ways to define similarities directly on this space [109]. Then the need of recovering correspondences is implicitly addressed. In such a context, well known metrics from the space of images have been adopted, like the sum of squared differences criterion, the normalized cross correlation, the mutual information, etc. Furthermore, numerous techniques can be used for optimizing these cost functions, like gradient descent, relation, combinatorial optimization, etc. Most of these methods are quite sensitive to the initial conditions and have no guarantee of convergence to the optimal solution if the deformation model is too dense (two many parameters to be determined).

The review of existing work in the area of shape registration shows that the representation problem is quite well addressed, which is also the case for the deformation models. On the other hand, defining appropriate metric functions that measures the similarities between shapes is still an open problem. The choice of a proper metric is related to every aspect of image registration, that are representation, registration model and similarity. However, it seems that the choice of implicit representations can provide enough freedom on exploring various metrics as well as various deformations models. In terms of optimization, all methods presented in the literature are to a certain degree sensitive to noise. Moreover when solving mathematical inference and imposing geometric consistency even the optimal solution to the problem cannot guarantee that the obtained correspondences are the right anatomical ones. One can overcome this limitation through means of providing quantitative and qualitative registration results. In such a context we view registration as a statistical estimation problem when one aims to recover the mean solution and the local uncertainty.

The remainder of this chapter is organized as follows. In section 2 we review shape representations, while in section 3 we discuss prior art on global and local registration. The last introductory section of this chapter describes the state of the art in optimization techniques. Then, we introduce our general registration approach that involves a global and a local component focusing on the in-

troduction of uncertainties. This model is customized to deal with organ-specific problems through the use of a TPS deformation model, as well as generic deformation models through an FFD.

2.2 Shape Definition and its Relation to Deformation

Contours, surfaces, boundaries and their mathematical descriptors often called shape are the key to our world as they are the most important visual features for the identification of most objects. In the scientific literature, various definitions of the term shape exist. All of them agree to limit the shape to the geometry, subtract any color or illumination information and therefore differentiate it from the 'appearance'. A simple definition of the space of surfaces and volumes considers a measurable subset of $\mathbb{R}^2$ or $\mathbb{R}^3$. However this is too generic, and not efficient from a computational perspective, therefore more specific subsets should be considered in computer vision. The work of Delfour and Zolesio [54] and Charpiat [27] add some constraints to this basic definition to create the set of 'shape of interest'. These definitions are implicitly used in most of the state of the art, but require an explicit and somehow restrictive form to prove the equivalence between common distance functions defined in this manifold.

Let $\mathcal{S}$ represent a measurable subset of $\mathbb{R}^2$, contained in a 'hold all' open bounded domain also called the 'image domain' Ω. It is possible from this point to define the set of smooth shapes $\mathcal{C}_0$ (resp. $\mathcal{C}_1, \mathcal{C}_2$). For any x on the boundary of $\mathcal{S}$ ($\partial\mathcal{S}$) there exist a neighborhood $\mathcal{N}$ and a C_0 (resp. C_1, C_2) scalar function (f) such that $\partial\mathcal{S} \cap \mathcal{N}$ can be represented as the epigraph of f:

$$\partial\Omega \cap \mathcal{N} = \{x \in \Omega | f(x) <= 0\}.$$

This definition implies that the boundary of a $\mathcal{C}_1$ or $\mathcal{C}_2$ cannot have multiple points and is represented by a simple regular curve (or surface in the 3D case). Restricting the shape of interest to the set of smooth shapes is usually considered in most of the literature as it is compatible with the discrete computer world. However considering smooth shapes in the continuous space $\mathbb{R}^n$ is not sufficient as one can design sequences of shapes with constant area and perimeter diverging to infinity. This kind of paradox was formally removed by introducing a constraint on the maximum curvature of the boundary $\partial\mathcal{S}$. This constraint may be released on a finite set of points to allow shapes with sharp angles. From this formal definition, one can extract two particular subsets *full shapes* and *boundary shapes*, which are a practical way to classify the descriptors:

- The full shapes $\mathcal{F}$ are defined as the subset of $\mathcal{S}$ that satisfy that the closure of their interior is themselves ($\mathcal{F} = \{\Omega \subset D :\overset{\circ}{=} \Omega\}$).

- The boundary shape $\mathcal{B}$ is the subset of $\mathcal{S}$ such that the shape is equal to its boundary ($\mathcal{B} = \{\Omega \subset D : \partial\Omega = \Omega\}$).

A slightly different approach to the concept of shape relates it directly with its object counterpart which can be presented under several poses and therefore present certain variability. Kendall in 1977 [94], and Dryden and Mardia in 1988 [112] proposed a simple shape representation that consisted of points sets and defined the resulting shape as whatever remains once a simple groups of transformations such as similarity (translation, rotation and scaling) is factored out. Therefore a shape can be defined as the equivalence classes of object views according to such group of transformations. However there is at least three reasons that make the task of identifying shape according to the former definition challenging: (i) 2D images are the representations of objects from the real 3D world, so the 2D problem would take advantage of considering the projective transform rather than similarity. Furthermore, occlusions and degenerative views (different 3D views producing the same 2D projections) make the shape definition non-unique.. (ii) Two shapes perturbed due to noise should present minor local differences and have to be identified as identical. This introduces robustness in the representation and the importance of it also depends on the application. (iii) Recovering the geometric properties of the shape (not only the representation) is critical since such properties are the core features to registration. This definition of shape plus deformation find its foundation in [193], where the process of identifying similar shapes using deformable grids was introduced. In such a way one can say that the concept of deformable template was born 90 years ago. This idea was then formalised in the 80's introducing the group action defining the variability within object class in the General Pattern Theory as introduced by Grenander in [75].

It is clear that many tasks in computer vision have to do with the comparison of shapes, having two different shapes, one should be able to say whether these come from the same object. Therefore we denote two categories of shape representations which are used in computer vision:

- *Moments* also known as *descriptors* transform a shape into a set of features or 1D functions. These features are often designed to carry invariant properties with respect to some group of spatial transformation and therefore offer some robustness in identification and shape comparison tasks.

- *Geometric Models* present a set of features or usually scalars or functional that are used to reconstruct a unique shape in a continuous manner. For identification purpose, these types of representation are closely related to shape registration. Geometric models are twofold:

  Discrete Geometric Models: A shape is represented with a finite set of scalars, a continuous shape may be reconstructed given a interpolation framework between the scalar infor-

mations.

Continuous Geometric Models: The shape interior or boundary is directly represented with the use of an explicit of implicit function. Continuity and smoothness are directly enclosed in the properties of the function.

2.2.1 Shape Representation with Moments

Shape moments generally look for effective and perceptually important features on either shape boundaries or inner regions defined by them. One of the most desirable property of a shape descriptor is that two different shapes should have distinct descriptors, and respectively perceptually similar shapes should have similar descriptors. These properties do depend upon the definition of shape (including or not the group invariance), and the amount of precision/discriminative power required by the application. Most often, once shape description is defined, the next task to be addressed is shape matching, ie. determine whether two objects belong to the same class. We refer to [215] for an extended review on shape descriptors.

Having the outline of a continuous shape defined (boundary shape), including the interior region of it one now can proceed to a finer classification of descriptors according to their definition domain. Several descriptors have been considered in the literature. Simple shape descriptors as area, circularity, convexity, principal axis ratio have been studied in [141].
A well known shape descriptor named *shape context* [10] for discretized contours builds for each of the N contour points, N histograms of the segments links to every other point on the contour, represented in polar coordinates. These histograms are concatenated in one large vector describing the shape in a translation and rotation invariant fashion. The shape context was generalized to surfaces in 3D, but 3D rotation invariance proved difficult to ensure and other approaches simply use a cylindrical parameterisation of the histograms [97].
Shape signature is a class of descriptors that represents a 2D shape with a 1D function. This can be centroidal profile, tangent angle, curvature, or chord length. However these methods are not efficient for matching because they involves 1D or 2D matching in the feature space. Consequently moments are used to reduce the dimensionality of the shape signature [170].
Another class of methods used to describe shapes relies on *linear scale space*: Asada and Brady [5] considered tracking inflection points on a parameterized curve if it is submitted to the heat equation. The shape becomes smoother and at certain particular scales, inflections points disappear one by one. This results in an *interval tree*, that is the diagram representing the curve abscissa of a disappearing inflection point, and the scale at which this inflection point disappears. This tree is used for matching and often combined with other curvature based information.

Region descriptors are based on full shapes, and therefore less sensitive to noise and local variations of the shape boundary. Computation of *geometric moments* has been very attractive because combination of low order moments leads to similarity invariant descriptors [11]. On the other hand this method considers few invariants to describe a shape and therefore is not precise and is too sensitive to small variations. Also *orthogonal moments* have been introduced, weighted with a set of orthogonal functions as Legendre, Zernike and pseudo Zernike polynomials. In [68] these moments are combined into similarity invariant descriptors at any order, and proved robust to noise [101]. More recently integral invariants signature were introduced which share an invariant property of the boundary curvature [110, 109]: by considering the surface (resp. volume) of the shape located within a certain radius of any boundary point, one can build a descriptor invariant under any rigid transformation. Varying the radius size also shows a multiscale approach that has been used for shape matching.
Different approaches to region based descriptors are defining the structure of the surface. One is based on the *convex hull*, defined as the smallest convex shape which encompasses the shape under study. The difference between the two convex hulls and the shape defines a set of smaller non-convex shapes. In an iterative manner, the convex hull of every shape is extracted, until the original shape can be written as the sums and differences of convex elements. This descriptor can be presented as a tree where every son node is a convex hull added or removed from the parent node [51, 170]. Another approach makes use of the medial axis or skeleton, defined as the loci of the centers of all the bitangent circles to the boundary of the shape [14]. This representation is transposed to a graph for matching purposes and extensive work has been carried on this representation both in 2D and 3D to improve noise robustness.

Moment-based shape representations are quite efficient for recognition purposes but is not well-adapted to registration applications. This is due to the fact that such models have limited representation of the local geometry. Discrete geometric models is the simplest possible representation which can account for the geometry.

2.2.2 Shape Representations with Discrete Geometric Models

Discrete representation allows to represent the evolution of shape directly with an explicit numerical scheme on the surface control points. Limiting the deformations to prevent topological problems is necessary and one has to add local regularization constraints. Discrete contour and triangulation methods represent the model as a set of points (vertices) with a neighborhood re-

lationship [206]. Also these deformable models are combined with a discrete scheme to append vertices to the triangulation [61] or update the topology via Delaunay based remeshing [144].
Delingette in 1994 introduced Simplex meshes [55], presenting a constant vertex connectivity: each vertex is connected to exactly 3 others. This model is extensively used in medical imaging for segmentation and tracking [120].
Physical models based on masses and springs were also proposed in [199, 145] to create adaptive meshes and in [125] where the spring model is combined with modal analysis to reduce the complexity. A different physics based deformable model relies on a set of particles evolving according to the Newton dynamics laws [179]. In order for the particles to model 3D surfaces, each particle represents a surface element and carry a normal vector. One has to design an internal energy between particles, favoring co-planarity and co-circularity that tends to organize them into regular surfaces.

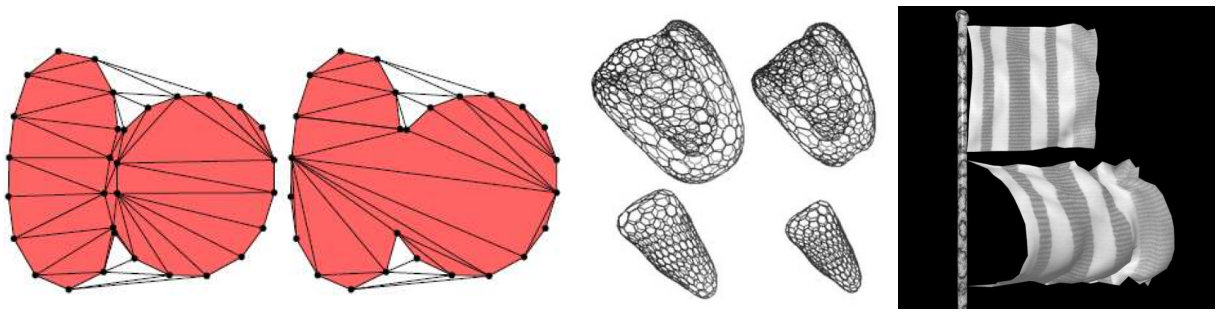


Fig. 2.1: From left to right, Delaunay Remeshing [144], Simplex representation [120], Mass Spring model [145].

Another kind of discrete models is creating a continuous and differentiable interpolation between control points. This is the case of a parametric contour using cubic B-spline interpolation between control points called B-snake [117], taking advantage of both local control and continuity. It may also handle sharp edges by increasing the number of control points. The method was also adapted to medical imaging for contour tracking [76]. Spline based methods have shown many evolution also for 3D surface representation with NURBS, commonly used in CAD applications [100] and surface extraction from images [192].
B-spline is just one kind of polynomial representation that is related to a general case following from the finite element theory: If contours or surfaces need to be expressed in an explicit way, and are related to an energy minimization scheme, it can be shown that using finite elements methods for the minimization can lead to different representations [37, 142]. Indeed, this discretization process requires the solution to be expressed in a finite dimensional vector space. Then, by choosing an appropriate base of piecewise polynomial functions with compact support satisfying continuity and differentiation conditions. Optimization of the geometric model with respect to its parameters

ends up in a linear system of the coefficients of these polynomials.

Seeking representations which can encode shape variability among training examples was the next challenge. The m-reps rely on a skeleton graph [143], extracted from the medial representation of a reference shape and subdivided into a regular lattice. The nodes of the medial representation are linked to the surface with a set of segments. Variations of the shape induced by the deformations of the skeleton and the continuous surface are reconstructed with a subdivision algorithm (Catmull-Clark).

Discrete models refer to a reasonable compromise betweeen complexity and geometric correctness. Their main limitation is the lack of explicit method to determine the local geometry of the shape and the associated geometric properties.

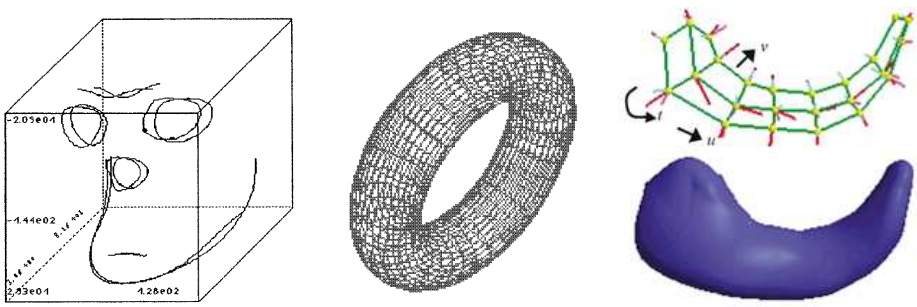


Fig. 2.2: Examples of Continuous Models representation. From Left to Right, B-spline curve fitting [76], Fourier surface Representation [171], m-reps [143]

2.2.3 Shape Representations with Continuous Geometric Models

Among all shapes representations, geometric models have probably been the most explored as their capability of evolving in time is particularly adapted to registration and segmentation purposes. Therefore in the literature, they are often bounded to an evolution model related to the particular application. This section however will strictly focus on the representations. Geometric models can be represented by the boundary of the shape of interest (either closed or open). The full shape can also be represented in an implicit manner and will be developed in the next section. A more extended review on geometric models can be found in [121].

Continuous models refer to shapes where geometric derivatives can be computed at any point on the boundary (contour or surface). Discretization occurs in a subsequent step for implementation purposes. The most widely referred deformable models were introduced by Terzopoulos and Kass around 1987 [189, 93]. The so called *snakes* simply use a geometric representation as a

parametric curve $\mathcal{C}$:

$$\mathcal{C} \ : \ s \in [0,1] \rightarrow (\mathbf{x}(s), \mathbf{y}(s))\mathbb{R}^2.$$

This curve was designed to minimize an energy functional $E(\mathcal{C})$ depending upon a regularity and a matching criteria. $\mathcal{C}$ can be transformed into an evolving contour by using gradient based minimization methods. The representation was rapidly adapted to 3D [190] with curves and surfaces:

$$\mathcal{A} \ : s \in [0,1] \rightarrow (\mathbf{x}(s), \mathbf{y}(s), \mathbf{z}(s))\mathbb{R}^3.$$
$$\mathcal{S} \ : (r,s) \in [0,1]^2 \rightarrow (\mathbf{x}(r,s), \mathbf{y}(r,s), \mathbf{z}(r,s))\mathbb{R}^3.$$

The representation was largely used for medical imaging purposes [115, 38]. One may notice that this representation requires additional conditions on the boundary of the parameters set defining the topology of the shape (for instance $\mathcal{C}(0) = \mathcal{C}(1)$ defining a circular contour). Snakes and deformable surfaces are consequently discretized on a regular grid and simply represented by a structured set of points. Discretization of the snake or deformable surface is natural when minimizing the energy functional using finite difference methods.

A different type of parametric representations does not rely on a structural decomposition of the contour or surface as previously but on global components acting on the entire model in a linear manner. The so-called modal decompositions are split in two categories:

- *Fixed modes* decomposition represents a closed contour decomposing its polar representation in Fourier basis [171], while evolution of the same idea made use of wavelets thus adding locality and scale [213]. The method was also extended to 3D surfaces using the spherical harmonic basis [178]. A different approach was introduced in [140] where 'free vibrating' modes of the contour were computed relying on a physical modeling of the surface.

- *Eigen modes* decomposition relies on a training sample. The representation directly depends on the application and is also limited to a linear combination of eigen-modes modeling contour deformations. This method was initiated by Cootes, starting in 1993 [43], using principal component analysis in combination with point distribution models.

While these geometric models are mostly defining the boundary shape and therefore evolution is always directly related to the parameterization, a different class of representation only requires discretization of the shape boundary for display purposes. These models are a trade-off between continuous and discrete models since the underlying representation depends on the discretization of the model. They fail to describe shapes with multiple non-connected components and require frequent reparametrization to determine the local geometric properties.

2.2.4 Implicit Representations

Implicit shape representation is actually one of the most natural as it directly refers to the definition of subset we have presented in section 2.2. Referring to a scalar function defined in Ω, it represents any shape $\mathcal{S}$ as:

$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) <= 0\}$$

This representation is also commonly used to represent closed contours or surfaces ($\mathcal{B}_f = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = 0\}$). This representation augments the representation space. In 3D for instance surfaces are a 2D manifold as the implicit representation is lying in a functional vector space. Therefore constraints are added to f to avoid computation on every point of the 3D space and regain part of the efficiency of the explicit models.

Algebraic surfaces is a class of implicit representation where f is a polynomial function. These representations have been used to reconstruct continuous surfaces from 3D unstructured sets of points [187]. The representation shows similarity invariants and interpretable coefficients [180]. Most of the theoretical work is oriented to design polynomials that preserve the stability of the shape with respect to the coefficients of the high order polynomials [186].

Superquadrics and hyperquadrics are two particular classes of algebraic surfaces that constrain the family of shapes represented. Superquadrics were used in medical imaging as their small number of parameters has shown a certain robustness to noisy data [8]. Also hyperquadrics are defined as an extension of superquadrics. The implicit equation is written as:

$$f_{\mathbf{p},d}(\mathbf{x}) = \sum_{j=1}^{n} \mid \mathbf{p}_j.\mathbf{x} + d_j \mid^{\epsilon_j} - 1 = 0,$$

where $\{\mathbf{p}_j\}$ is a set of vectors, $\{d_j, \epsilon_j\}$ a set of scalars with $\epsilon_j > 0$ and n an integer larger than 3. These shapes are bounded and convex. Additional terms were added to the implicit form in order to represent concavities [35].

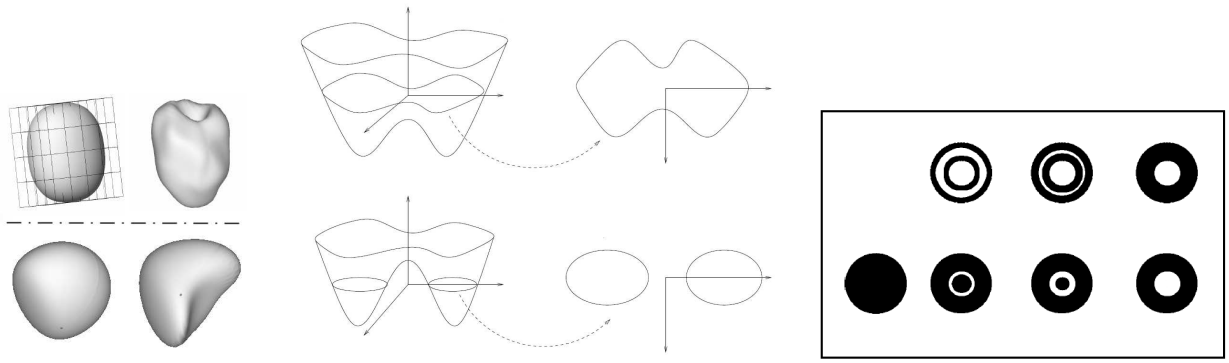
Shape Representation based on kernels is using for f a finite dimensional Reproducing Kernel Hilbert Space. It is constrained to a general expression with the form:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{p}_i, \mathbf{x}),$$

with K being the reproducing kernel. It was used in [3] with an elegant statistical continuous framework modeling both template and deformations with Gaussian kernels. This approach to

implicit representation was also frequently used to infer a full shape only based on a finite sampling of points in space [204]. This approach is therefore closely related to radial basis function interpolation that will be further explained in the next section.

These parametric approaches can be limited to describe more complex structures than closed contours and surfaces. A general implicit representation is based on Euclidean distance map and may represent visual structures of any dimension.



(a) (b) (c)

Fig. 2.3: Implicit representation: (a-top) superquadrics [8] (a-bottom) hyperquadrics [35], (b) topology change on level-set representation, (c) phase field toppology change [148].

Distance Transform

The signed distance function of any shape lying in the image domain is expressed with the following expression:

$$\phi_{\mathcal{S}}(\mathbf{x}) = \begin{cases} d_{\Omega}(\mathbf{x}) & = ED(\mathbf{x},\mathcal{S}) = \inf_{y\in\mathcal{S}} \|\mathbf{y}-\mathbf{x}\|, & \mathbf{x}\in\mathcal{CS} \\ -d_{\mathcal{CS}}(\mathbf{x}) & = -ED(\mathbf{x},\mathcal{CS}) = -\inf_{y\in\mathcal{CS}} \|\mathbf{y}-\mathbf{x}\|, & \mathbf{x}\in\mathcal{S} \end{cases} \tag{2.1}$$

$$= \chi_{\mathcal{S}}(\mathbf{x}).ED(\mathbf{x},\mathcal{S}) - (1-\chi_{\mathcal{S}}(\mathbf{x})).ED(\mathbf{x},\mathcal{CS}) \tag{2.2}$$

where $\mathcal{CS}$ represent the complement of $\mathcal{S}$ in the image domain Ω, ED the Euclidean point to shape distance and $\chi_{\mathcal{S}}$ the characteristic function of the shape $\mathcal{S}$.

This representation is also the cement of a surface evolution technique named Level sets. Level sets have been widely used as an alternative to explicit contour and surface representations as they are based on an equivalence between the motion of a contour along its normal and the global deformation of an associated implicit deformation. Introduced by Dervieux and Thomasset in the field of fluid dynamics [56, 57], it first came to computer vision with the work of Osher and Sethian

[127]. Introduced in medical imaging by Malladi in [108] this technique lead to many different applications [126]. Level set techniques do not require parameterization and allows topological changes of the shape. Many evolutions were carried to improve speed and preserve distance constraints during evolution. The distance is a Lipschitz function with coefficient 1 and its gradient is defined almost everywhere except on the skeleton and has unit norm. This representation can be used to represent any mixture of full and boundary shapes. In the case of a shape lying in the boundary set $\mathcal{B}$, the distance is always positive and its gradient is also undefined on the shape itself.

The representation is invariant under a rigid transformation and can be slightly modified to account for a similarity transform. Consider $\mathcal{L} = (s, R, T)$ a similarity transform, Let the shape $\mathcal{S}'$ be the transformation by $\mathcal{L}$ of the shape $\mathcal{S}$, then the relation $\phi_{\mathcal{S}}(\mathbf{x}) = \phi_{\mathcal{S}'}(\mathcal{L}(\mathbf{x}))$ holds everywhere, and can be proven:

$$\begin{aligned} ED(\mathbf{x}', \mathcal{S}') &= \inf_{\mathbf{y}' \in \mathcal{S}'} \|\mathbf{y}' - \mathbf{x}'\| \\ &= \inf_{\mathbf{y}' \in \mathcal{S}'} \|\mathbf{y}' - (s.R.\mathbf{x} + T)\| \\ &= \inf_{\mathcal{L}^{-1}(\mathbf{y}') in \mathcal{L}^{-1}(\mathcal{S}')} \|\mathbf{y}' - (s.R.\mathbf{x} + T)\| \\ &= \inf_{\mathbf{y} \in \mathcal{S}} \|s.R(\mathbf{x} - \mathbf{y})\| \\ &= s. \inf_{\mathbf{y} \in \mathcal{S}} \|\mathbf{x} - \mathbf{y}\| \\ &= s.ED(\mathbf{x}, \mathcal{S}) \end{aligned}$$

This proves that the representation is invariant under rotation and translation and can be easily modified to account for scaling factor. If one calls $\mathcal{DT}$ the manifold of distance transforms, and augments this space by including scaling, then $\mathcal{DT}' = \{s\phi | \phi \in \mathcal{DT} \ \& \ s \in \mathbb{R}\}$. Considering the equivalence relation on the set $\mathcal{DS}' : \phi_{\mathcal{S}_1} \sim \phi_{\mathcal{S}_2} \leftrightarrow \exists s \in \mathbb{R}, \phi_{\mathcal{S}_1} = s.\phi_{\mathcal{S}_2}$, such representation is invariant under similarity when considered in the quotient space $\mathcal{DT}'/\mathbb{R}$. The stability of the distance with respect to small perturbation has been shown in [220]. Indeed small variations of a shape within a certain radius induce bounded variation of the sum squared difference of the distance transforms, and therefore proves that a registration problem stated using distance transform is well-posed.

Also this representation proved efficient for shape matching using calculus of variations as the non-zero iso-level provides additional support for optimization. One should note that the non-zero level set provides a simpler representation of the shape: The $r > 0$ level sets correspond to the boundary of the shape dilated with a ball with radius r. Therefore any concavity of the shape with curvature smaller than r are smoothed out (Fig. 2.4).

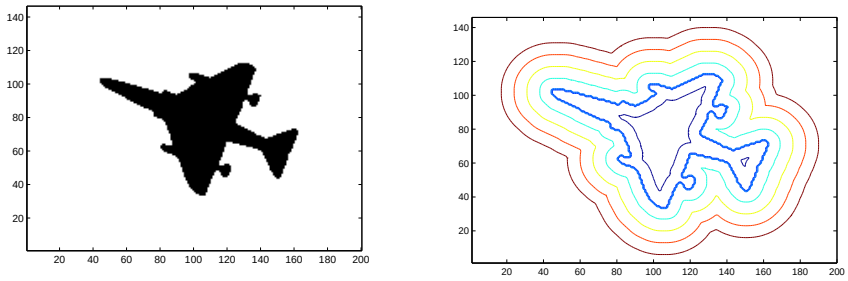


Fig. 2.4: Displaying zero and non-zero iso-contour of the distance transform extracted from a sample shape of plane.

Most of the representation techniques presented above were associated to evolution schemes able to perform either registration or deformation. A different approach would consider the representation as being fixed frame and applying a subsequent regular transformation. This is the key idea of a deformable template, and therefore we need to review part of the state of the art on classical transformations.

2.3 Spatial Transformations

As previously stated the shape registration process is directly related to the choice of a transformation that is usually unrelated to the choice of the representation considered for the shapes. We will now introducing notations that will be used in the remaining of the chapter. Let us consider $\mathcal{S}$ a source shape that is to be transformed, and $\mathcal{T}$ the target shape onto which $\mathcal{S}$ has to be deformed and aligned. We will denote by $\mathcal{L}$ a transformation on the entire domain Ω and restrict ourselves to parametric deformation so that $\mathbf{\Theta}$ refers to a vector of scalar parameters:

$$\begin{array}{rccl} \mathcal{L}: & \mathbb{R}^n \times \Omega & \rightarrow & \Omega \\ & (\mathbf{\Theta}, \mathbf{x}) & \mapsto & \mathbf{y} = \mathcal{L}(\mathbf{\Theta}, \mathbf{x}) \end{array} \tag{2.3}$$

Then the transformation of $\mathcal{S}$ under $\mathcal{L}(\mathbf{\Theta}, .)$ can be considered and will be refered as $\mathcal{S}'$

$$\mathcal{S}' = \mathcal{L}(\mathbf{\Theta}, \mathcal{S}) \Leftrightarrow \mathcal{S}' = \{\mathbf{x}' = \mathcal{L}(\mathbf{\Theta}, \mathbf{x}) | \mathbf{x} \in \mathcal{S}\}$$

Hence if we denote by $\mathcal{R}$ the shape representation process, and E_{data} the similarity/dissimilarity metric defined in the space of representations, the registration process can be simply rewritten as:

$$\mathbf{\Theta}_{min} = \arg\min_{\mathbf{\Theta}} E_{data}(\mathcal{R}(\mathcal{L}(\mathbf{\Theta}, \mathcal{S})), \mathcal{R}(\mathcal{T})) \tag{2.4}$$

The scope of a spatial transformation refers to the influence of the elements of the vector $\mathbf{\Theta}$. A transformation is called global if the whole domain is transformed when modifying any parameter of $\mathbf{\Theta}$. On the other hand it will be considered as local if the modification of a parameter affects the domain locally. In this latter case, the vector of parameters will be associated to a set of particular points of the domain.

2.3.1 Global Deformations

Global deformation may only be used to change the pose of a shape. In the context of shape registration it performs a rough registration as a necessary step before considering local deformations.

The simplest transformation is rigid, also called Euclidean transformation and can be expressed in the two dimensional case with the expression:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix} . \begin{bmatrix} x - C_x \\ y - C_y \end{bmatrix} + \begin{bmatrix} tx + C_x \\ ty + C_y \end{bmatrix},$$

where (C_x, C_y) represents a fixed point chosen as the center of rotation, θ is the angle of rotation and (tx, ty) the translation vector. Choosing the center of the object of interest for (C_x, C_y) is essential for the convergence of the registration process. The degree of freedom of a linear 2D transformation can be augmented to obtain separately the 6 degrees of freedom of an affine transformation. However we use a specific form that allows to control separately the physical deformation of the domain:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s. \begin{bmatrix} a & b \\ 0 & 1/a \end{bmatrix} . \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix} . \begin{bmatrix} x' - C_x \\ y' - C_y \end{bmatrix} + \begin{bmatrix} tx + C_x \\ ty + C_y \end{bmatrix},$$

where s is the scaling factor and represents the global change of volume, a the anisotropic dilations along the two axes and b the shearing factor.

In the 3D case rotation could be explained with the composition of 3 successive and orthogonal planar rotations (through the Euler angles). However the resulting expression is rather expensive and the three angles are inter-dependant. Therefore a vectorial expression using the exponential form is preferred. Considering the vector $\omega = \theta.\omega_0 = (rx.\theta, ry.\theta, rz.\theta)$ representing the rotation of angle θ around axis (rx, ry, rz) with unit norm, the rotation matrix associated with the vector ω

is:

$$R_\omega = \exp\left(\theta.\begin{bmatrix} 0 & -rz & ry \\ rz & 0 & -rx \\ -ry & rx & 0 \end{bmatrix}\right) = \exp(\theta.\hat{\omega_0}) = \exp(\hat{\omega})$$

Developing this expression with Taylor expansion leads to a matrix form of the Rodrigues's fomula:

$$R_\omega = I + \hat{\omega_0}.\sin(\theta) + \hat{\omega_0}^2.(\cos(\theta) - 1)$$

Finally transforming this expression simply using the rotation vector ω:

$$R_\omega = I + \hat{\omega}.\frac{\sin(\|\omega\|)}{\|\omega\|} + \hat{\omega}^2.\frac{\cos(\|\omega\|) - 1}{\|\omega\|^2}$$

Computation of derivative on this expression is more efficient and does not favor any direction. An expression including rotation center, scaling and translation is used to represent a $3D$ similarity transform ($\mathbf{x}' = sR(\mathbf{x} - \mathbf{c}) + \mathbf{T} + \mathbf{c}$).

One may also mention perspective transform. Such transformation makes use of the homogeneous coordinates, that is obtained embedding the d-dimensional space in a d dimensional manifold called projective cone and defined as follows in the $2D$ case [80]:

$$[x, y, \omega] \sim [x', y', \omega'] \Leftrightarrow \exists\lambda \in \mathbb{R} | x' = \lambda.x, y' = \lambda.y, \omega' = \lambda.\omega$$

Then any $2D$ point (x, y) of the space is embedded into the projective space with homogeneous coordinates $[x, y, 1]$. Using these coordinates allow to define an homography (Fig. 2.5), that is a perspective transform describing the change in aspect of a planar object under different camera views. Considering such transformation can be of primary interest when the registration application is related to image views of objects like surveillance. It can be expressed as a simple linear transform in the projective space:

$$\begin{bmatrix} x' \\ y' \\ \omega' \end{bmatrix} = A_{3,3}.\begin{bmatrix} x \\ y \\ \omega \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{bmatrix}.\begin{bmatrix} x \\ y \\ \omega \end{bmatrix}$$

A few more complex global transformations have been considered in the literature. A quadratic transform for instance was used for retinal image registration, as a fair approximation of the deformation of the curved retinal surface under different views [21].

The estimation of the parameters of any global registration problem is using all the features

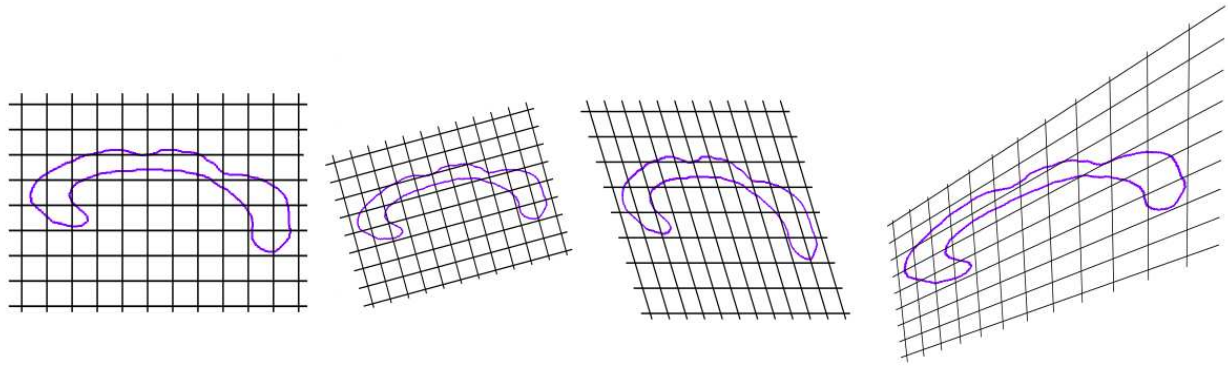


Fig. 2.5: Classic global transformations. From left to right, identity, rotation+scaling, affine, perspective transformation

of the shape representation which makes it an overconstrained problem. Therefore the problem of global registration is often well posed although optimization is not trivial. In particular uniqueness of the solution is not guaranteed and the performance heavily depends on whether or not the considered transformation can express the relationship between the two examples. Also the match between the transformed source shape $\mathcal{S}'$ and the target shape $\mathcal{T}$ is rough and does not allow to create one to one correspondence between shapes. Recovering local correspondences is an important task in shape analysis with application to numerous domains like biometric or medical imaging. Therefore we introduce a set of local deformations able to perform a better approximation towards shape warping.

2.3.2 Local Deformations

The simplest way to model local deformations is based on optical flow used for image registration [81]. It defines a continuous field of displacement (u, v) vectors on the whole image. Optical flow is a very well known ill posed problem: due to the aperture problem, only the normal component of the local displacements can be determined accurately. Therefore an additional constraint is imposed that forces smooth variation in the flow across the image and is used to constrain the tangential component of the displacement vectors:

$$E_{smooth}(\mathcal{L}) = \int_{\Omega} \|J_{\mathcal{L}}(\mathbf{x}) - I\|_F^2 \, d\mathbf{x} = \iint \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 dxdy,$$

where $\mathcal{L}$ is the transformation associated to the deformation field, $J_{\mathcal{L}}$ its Jacobian matrix and $\|;\|_F$ the Froebenius norm. Continuous deformations were also used for shape registration in [135] along with the same smoothness term. However such a large deformation field proved very expensive to

optimize, and difficult to regularize. Indeed, smoothness constraints need to be added on the whole image domain to preserve an invertible transformation and therefore preserve the topology of the transformed shape.

Many other popular non-rigid deformation are parametric and based on a set of control points. A control point is a fixed point of the domain whose displacement will induce a local deformation of the domain. These points shall be located either on a regular grid or scattered in the whole domain. These transformations actually find their foundations in the regularization and neural network theory [71]. The aim of this theory is to approximate an unknown function (in our case, deformation) with the knowledge of its value on a finite set of points in a d-dimensional space. The following framework was initially presented for the approximation of scalar functions. It can easily be extended to vector functions taking values in the same d-dimensional space, therefore defining a spatial deformation. Consider a set of control points $\{P_i\}_{i=1}^n$, with associated deformed point $\{P_i'\}_{i=1}^n$. Let $\mathcal{L}$ be the approximation of the unknown deformation field. This problem is underconstrained and a unique deformation $\mathcal{L}$ may be defined on the entire domain with the use of a smoothness or regularizing functional criterion $E_{smooth}(\mathcal{L})$. $\mathcal{L}$ is defined with the minimization of the functional:

$$H(\mathcal{L}) = \sum_{i=1}^{N} \|\mathcal{L}(P_i) - P_i'\|_2^2 + \lambda E_{smooth}(\mathcal{L})$$

The smoothness term refers to the oscillatory behavior of the function. Therefore the E_{Smooth} term will take a higher value if the deformation $\mathcal{L}$ shows high frequency when expressed in the Fourier domain. A common choice for the smoothness term therefore refers to the energy of $\mathcal{L}$ after the application of a high pass filter:

$$E_{smooth}(\mathcal{L}) = \int_{\mathbb{R}^d} d\mathbf{s} \frac{\|\tilde{\mathcal{L}}(\mathbf{s})\|_2^2}{\tilde{G}(\mathbf{s})}$$

where $\tilde{\mathcal{L}}$ and $\tilde{G}$ indicates the Fourier transform of $\mathcal{L}$ and G. $\tilde{G}$ is a symmetric positive function that tends to zero as $\|s\|$ goes to infinity. Using such a form for the smoothness and matching term, allows to separate the output components of the deformation $\mathcal{L}$. This proves that the vectorial problem stated here is a direct consequence of the scalar approximation problem commonly treated in the literature of regularization theory. This case is of particular interest as the optimum of the functional admits an exact solution of the form [106]:

$$\mathcal{L}(\mathbf{x}) = \sum_{i=1}^{N} \mathbf{w}_i G(\mathbf{x} - P_i) + \sum_{\alpha=1}^{k} \mathbf{d}_\alpha \psi_\alpha(\mathbf{x})$$

where $\{\psi_\alpha\}_{\alpha=1}^{k}$ is a basis of the k dimensional null space of the smoothing term and $(\{P_i\}_{i=1}^{N}, \{\mathbf{d}_\alpha\}_{\alpha=1}^{k})$ are two sets of d-dimensional vectors defining the transformation. Using a matrix form, one can express the unknown parameters $(\{P_i\}_{i=1}^{N}, \{\mathbf{d}_\alpha\}_{\alpha=1}^{k})$ of the transformation as follows:

$$\begin{array}{c}(G+\lambda I)\mathbf{w}+\psi^T\mathbf{d}=P' \\ \psi P=0\end{array} \tag{2.5}$$

where I is the identity matrix and other terms are matrices defined by:

$$\begin{array}{c}(P')_{ij}=(P'_i)_j,\ \ (\mathbf{w})_{ij}=(\mathbf{w}_i)_j,\ \ (\mathbf{d})_{\alpha j}=(\mathbf{d}_\alpha)_j \\ (G)_{ij}=G(P_i-P_j),\ \ (\psi)_{\alpha j}=\psi_\alpha(P_i),\end{array}_{c}$$

In the limit case where λ tends to zero, the match is exact ($\mathcal{L}(P_i)=P'_i$) and the deformation $\mathcal{L}$ is then called an interpolating function. Very different forms have been proposed for the smoothing term, leading to gridded or scattered spatial deformations. Some of them are object dependent while more generic ones are object dependent. In the scope of our work we have investigated both.

Free Form Deformations

This technique explicitly states the continuity and locality of the deformation. Free Form Deformation (FFD) relies on a set of control points which have to be located on a regular rectangular lattice. Spatial deformation is expressed as the tensorial product of deformation along each orthogonal direction. It was introduced in 1986 in [162] to deform computer graphic models. Consequently it was also used in the field of medical imaging [157, 20] for multimodal image registration.

Let us first introduce its relation to the general approximation theory and consider a regularization function that can be written in the frequency domain as a tensor product along different coordinates. As a consequence, regularization will be :

$$\tilde{G}(\mathbf{s})=\Pi_{j=1}^{d}\tilde{g}(s_i)$$

Applying the inverse Fourier transform, leads in the spatial domain to a basis function of the form:

$$G(\mathbf{x})=\Pi_{j=1}^{d}g(x_i)$$

Also the solution of the regularization problem in $\mathbb{R}^d$ can be decomposed independently with d regularization problems in $\mathbb{R}$.

However the true essence of FFD does not lie in the regularization as the form of the basis function will depend on the spacing between control points. Still the expression of the deformation will be expressed in a linear form. The two principal interests of FFD reside in the design of the basis function, not directly related to its Fourier form, it is built using a piecewise polynomial function with compact support and $\mathcal{C}^2$ continuity.

Consider a finite sequence of points in $\mathbb{R}$: $S = (x_1, x_2, \ldots, x_N)$ regularly spaced with distance l_S ($x_i - x_{i-1} = l_s$). Then a regular grid in $\mathbb{R}^d$ is built as the direct product of d sets of points S_i regularly spaced along each dimension, each of them containing N_i points spaced with distance l_{S_i}. Consider the multi-index $\alpha = (\alpha_1, \ldots \alpha_d)$ with $0 \leq \alpha_i < N_i$. The initial d-dimensional grid of control points is represented with the set of $\{P_\alpha\}$ and their new position as $\{P'_\alpha\}$. Then the deformation of any spatial point will be decomposed along each direction and expressed linearly with respect to the position of the points of P'

$$\mathcal{L}(\mathbf{x}) = \sum_\alpha \left(\Pi_{i=i}^d \chi_{S_i,\alpha_i}(\mathbf{x}_i)\right) P'_\alpha \,, \tag{2.6}$$

summing on all multi-index α where $\chi_{S_i,j}$ refers to the j^{th} B-spline basis function for the sequence S_i.

A B-spline basis of order n is a set of polynomials functions verifying $\sum_j \chi_{S,j} = 1$ with continuity $\mathcal{C}^{n-1}$. These basis function are defined recursively on the order n with the Cox-de Boor recursion formula. It was built as an extension to the Bezier Spline, based on Bernstein polynomials [53], as a B-spline has by construction a compact support. The most commonly used are cubic-B-Spline ($n = 3$) also in use to compute FFD, and shows $\mathcal{C}^2$-continuity everywhere.

$$\chi_{S,j}(x) = \begin{cases} \frac{1}{6}(u)^3 & \text{with } u = (x - x_{j-2})/l_S \quad \text{if } x \in [x_{j-2}; x_{j-1}) \\ \frac{1}{6}(-3u^3 + 3u^2 + 3u + 1) & \text{with } u = (x - x_{j-1})/l_S \quad \text{if } x \in [x_{j-1}; x_j) \\ \frac{1}{6}(3u^3 - 6u^2 + 4) & \text{with } u = (x - x_j)/l_S \quad \text{if } x \in [x_j; x_{j+1}) \\ \frac{1}{6}(1 - u)^3 & \text{with } u = (x - x_{j+1})/l_S \quad \text{if } x \in [x_{j+1}; x_{j+2}) \\ 0 & \text{elsewhere} \end{cases} \tag{2.7}$$

One can see from this basis function that $\chi_{S,j}$ is non-zero only in the space separating 4 control points. Therefore, at most 4 basis functions will be non-zero for any point in $\mathbb{R}$. As a consequence in the d-dimensional space, the deformation of any point $\mathbf{x}$ in space will only depend on the displacement of its 4^d neighboring points.

A Free Form Deformation is not an interpolation scheme as it does not verify $\mathcal{L}(P_\alpha) = P'_\alpha$. However optimization can be performed with respect to the position of the displaced control points.

Setting the generic vector of parameters with the position of control points : $\mathbf{\Theta} = \{P'_\alpha, \alpha_i = 1..N_i\}$ in a $\left(\Pi_{i=i}^{d} N_i \times d\right)$, the transformation in eq.2.6 can be written in a rather compact form:

$$\mathcal{L}(\mathbf{\Theta}, \mathbf{x}) = \mathbf{x} + \mathcal{X}(\mathbf{x}).\mathbf{\Theta}, \tag{2.8}$$

with $\mathcal{X}(\mathbf{x})$ a $(1 \times \Pi_{i=i}^{d} N_i)$ matrix depending non linearly on the location $\mathbf{x}$ and the initial position of the control points.

As registration is performed with a variational framework, the minimization in equation 2.4 is leading to an evolution scheme for the vector $\mathbf{\Theta}$ converging to the minimum of the functional. Using FFD, one cannot guarantee that the resulting transformation is invertible and therefore satisfies the desirable diffeomorphic constraint. To prevent such unwanted transformation to appear both an additional smoothing constraint, along with a coarse to fine approach on the number of control points in the grid have been introduced. Smoothing constraint is an energy term which increases singularly as the transformation is less regular. Ideally, this energy term should diverge to infinity as the associated transformation is not invertible. Defining a proper criterion for regularity that can be expressed in the parameters of the transformation is not straightforward. Therefore builds a criterion that will express the distance of the current transformation to a class of basic transformations associated to a 0 cost. We have added a second order smoothness constraint on the whole deformation that favors affine transformations. Its expression on the domain Ω can be expressed with the Froebenius norm of the Hessian matrix of the transformation:

$$E_{smooth}(\mathcal{L}) = \int_\Omega \sum_{\mu=1}^{d} \|\mathcal{H}(\mathcal{L}_\mu(\mathbf{x}))\|_F^2 d\mathbf{x},$$

where $\mathcal{L}_\mu$ is the μ-th component of the transformation $\mathcal{L}$. This expression is exactly the Thin-Plate-Spline energy that will be explored in the next section for scattered data interpolation. Restricting the domain Ω to the portion of the space covered by the grid of control points, given the polynomial form of the basis function, the substitution of equation 2.8 in this expression leads to a closed form expression acting directly on the displacement of the control points as a quadratic form:

$$E_{smooth}(\mathbf{\Theta}) = \mathbf{\Theta} C \mathbf{\Theta},$$

where C is a symmetric positive matrix. Consequently the optimization criterion is including both a similarity and a smoothing term, equation 2.4 can be rewritten:

$$\mathbf{\Theta}_{min} = \arg\min_{\mathbf{\Theta}} (1-\beta) E_{data}(\mathcal{L}(\mathbf{\Theta}, \mathcal{S}), \mathcal{T}) + \beta E_{smooth}(\mathbf{\Theta}),$$

where β is a positive constant that balances the contribution of the two terms ($0 \leq \beta \leq 1$). Given the form of the smoothing term and its local influence on the whole transformation, its value will be chosen according to the refinement of the grid. A larger value of β will be preferred at finer scale to prevent the apparition of local folding.

Refining Free-Form Deformations has been explored previously. In [162], different FFD were applied in a piecewise manner with adjacent grids, adding particular continuity constraints on the boundaries of the grid. Also refinements methods have been considered in the field of medical images [210], where FFD grids with different scale are superimposed at different locations, allowing a finer registration in certain areas. However the global deformation is explained as a sum of all the transformations at different scales. Our framework dedicated to shape learning, required the explanation of the whole transformation on a single grid. Therefore we have used a different approach capable of refining the grid on the whole domain. The solution to the refinement problem is not exact and requires an approximation technique to retrieve the best transformation at finer scale as the best estimate of the transformation at coarser scale. Consider a set of points located on the boundary of the shape to be registered ($\mathbf{P}_{\mathcal{S},i}$), a transformation at a coarse scale $\mathcal{L}_1$ and a transformation parameter vector $\mathbf{\Theta}_1$. Consider a transformation at a finer scale $\mathcal{L}_2$, we try to retrieve the best parameter vector $\mathbf{\Theta}_2$ minimizing the quadratic criterion:

$$\mathbf{\Theta}_2 = \arg\min_{\mathbf{\Theta}} \|\mathcal{L}_1(\mathbf{\Theta}_1, \mathbf{P}_{\mathcal{S},i}) - \mathcal{L}_2(\mathbf{\Theta}, \mathbf{P}_{\mathcal{S},i})\|^2$$

The transformation being linear in the parameter $\mathbf{\Theta}$, this problem admits a closed form solution.

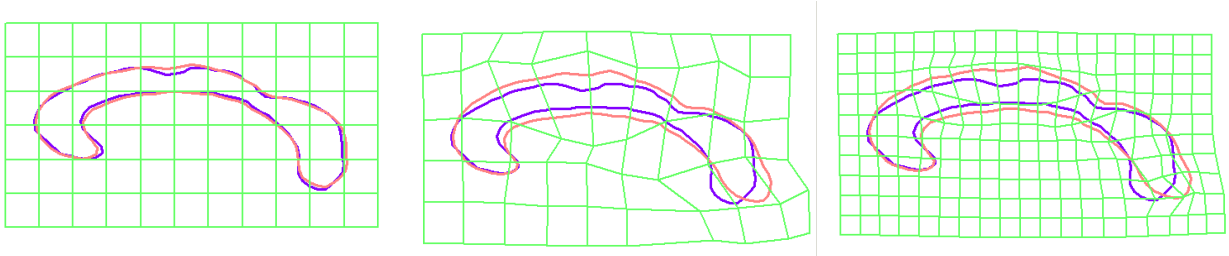


Fig. 2.6: Example of Free Form Deformations. (middle) deformations using a 6×10. (right) grid refinement with the same transformation and a 11×20 grid.

The principal drawbacks of free form deformations lie in the constraint on the location of the control points:

- Using a square grid to model deformation of non-square shapes gives very different importance to the control points. Control points located far from the shape boundary have very little influence and therefore may show a very high variability in their final position.

- Using control points located on a square grid with sufficient control on the whole shape leads to a high number of points especially if registration is performed in dimension 3.

It was necessary to adopt a different model for the local deformations that gives equal importance to the control points while using a smaller number of points to reach similar precision of the registration.

Thin Plate Splines

Thin Plate Spline (TPS) is a direct application of the approximation framework earlier introduced. A popular kind of smoothness function is invariant under a rotation ($E_{smooth}(\mathcal{L}) = E_{smooth}(\mathcal{L} \circ R)$), and corresponds to radial basis functions (RBF) of the form $G(\|\mathbf{x}\|)$. Many different kinds of RBF coexists and have been used for data interpolation such as Gaussian ($G(\mathbf{x}) = e^{-\beta\|\mathbf{x}\|^2}$) , multiquadric ($G(\mathbf{x}) = \sqrt{\|\mathbf{x}\|^2 + c^2}$), inverse multiquadric, etc...

The most widely used was developed by Duchon [62] and considers the smoothness term:

$$E_{smooth}(\mathcal{L}) = \int_{\mathbb{R}^d} d\mathbf{s} \|\mathbf{s}\|_2^{2m} \|\tilde{\mathcal{L}}(\mathbf{s})\|_2^2 \tag{2.9}$$

We will focus on the interpolation case ($\lambda = 0$) with $m = 2$ also known as Thin Plate Spline because it has a physical meaning and corresponds to the deformation of a thin sheet of metal (uniform elastic material) subject to a set of orthogonal displacement. This type of smoothness constraint was first introduced to model deformations in computer vision by Bookstein in [15] and applied to the warping of sets of points with an increasing number of control points. The thin plate spline smoothness term can be expressed in the spatial domain with the Froebenius norm of the Hessian matrix of the transformation:

$$E_{smooth}(\mathcal{L}) = \int_{\mathbb{R}^d} \sum_{\mu=1}^{d} \|\mathcal{H}(\mathcal{L}_\mu(\mathbf{x}))\|_F^2 d\mathbf{x} = \int_{\mathbb{R}^d} \sum_{\mu=1}^{d} \mathrm{tr}(\mathcal{H}(\mathcal{L}_\mu(\mathbf{x}))\mathcal{H}(\mathcal{L}_\mu(\mathbf{x}))^T) d\mathbf{x},$$

The null space of the smoothness terms corresponds to the polynomials of degree one, that is affine transformation. Therefore the solution of the interpolation problem can be explained with the form:

$$\mathbf{x}' = \mathcal{L}(A, T, V_i, \mathbf{x}) = A.\mathbf{x} + T + \sum_{i=1}^{N} \mathbf{w}_i G(\|P_i - \mathbf{x}\|), \tag{2.10}$$

where $A.\mathbf{x} + T$ represents the affine part of the transformation and the set of vectors $\{V_i\}_{i=1}^n$ the

weights of the non affine warping. The associated radial basis function $G(\mathbf{x})$ admits an explicit form depending upon the dimension of the problem:

$$G(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|^2 \log(\|\mathbf{x}\|^2) & \text{d = 2} \\ \|\mathbf{x}\| & \text{d = 3} \end{cases}$$

Following equation 2.5 with $\lambda = 0$, the unknown parameters are the solution of the following linear problem:

$$\begin{bmatrix} G & P^T & \mathbf{1}_N \\ P & & \\ {\mathbf{1}_N}^T & \mathbf{0}_{d+1,d+1} & \end{bmatrix} \begin{bmatrix} W^T \\ A^T \\ T^T \end{bmatrix} = \begin{bmatrix} P'^T \\ \mathbf{0}_{d,d} \\ \mathbf{0}_{1,d} \end{bmatrix}, \tag{2.11}$$

where $G_{i,j} = U(\|P_i - P_j\|)$, $P = (P_1, \dots P_N), W = (\mathbf{w}_1, \dots \mathbf{w}_n)$, and $\mathbf{1}_n, \mathbf{0}_3$ are columns vector of 1 and 0. **Let L be the $((N + d + 1) \times (N + d + 1))$ matrix in the left part of the equation.**

In the present case we are interested in transforming any 3D point $\mathbf{x}$ into $\mathbf{x}'$ with respect to the displacement of the control points. This is known as the dual form of the interpolation problem. Expressing equation 2.10 in its matrix form:

$$\mathbf{x}' \;=\; [W, A, T].[B, \mathbf{x}^T, 1]^T,$$

where $B = (B_1, \dots B_N)$ and $B_j = U(\|P_j - \mathbf{x}\|)$. Combining this equation with equation 2.11 writes the displacement of any point as the linear combination of the new position of control points:

$$\mathbf{x}' \;=\; [P', 0_{d,d+1}].L^{-1}.[B, \mathbf{x}^T, 1]^T.$$

Therefore the deformation of any $\mathbf{x}$ point can be expressed as a linear combination of the displacement of the control points. The non linear part only depends upon the initial position of $\mathbf{x}$ and control points $\{P_i\}$.

Similar to FFD, the TPS transformation is optimized in a variational framework on the position of the control points and therefore does not guarantee an invertible transformation. We adopt both a regularization and coarse to fine strategy to address this issue.

Regularization is based on the smoothing term of the thin plate spline $E_{smooth}(\mathcal{L})$. From the proof on general form of solution of the regularization problem [71], one can easily show that this energy is directly related to the position of the control points:

$$E_{smoooth}(\mathbf{\Theta}) = \|W.G.W^T\|_F^2 = \|P'.L_N.P'^T\|_F^2,$$

where L_N is the submatrix of L, formed by the rows and columns $(1..N)$. Consequently the optimization criterion is including both a similarity and a smoothing term, equation 2.4 can be rewritten as:

$$\mathbf{\Theta}_{min} = \arg\min_{\mathbf{\Theta}}(1-\beta)E_{data}(\mathcal{L}(\mathbf{\Theta},\mathcal{S}),\mathcal{T}) + \beta E_{smooth}(\mathbf{\Theta}),$$

where β is a positive constant that balances the contribution of the two terms ($0 \leq \beta \leq 1$). Given the form of the smoothing term and its local influence on the whole transformation, its value will be chosen according to the number of control points. Larger value of β will be preferred when increasing the locality of the transformation.

In the case of TPS, as the transformation is an interpolation scheme, going from a scale to the next is simply performed by applying the previously retrieved transformation to the larger set of new control points. Formally, consider two sets of control points used for two successive scales $S_1 = \{P_i, i = 1..N_1\}$ and $S_2 = \{P_i, i = 1..N_2\}$ with $N_1 > N_2$. Suppose that the optimum transformation has been obtained for the set S_1:

$$\mathbf{\Theta}_{1,min} = \arg\min_{\mathbf{\Theta}} E_{data}(\mathcal{L}_1(\mathbf{\Theta},\mathcal{S}),\mathcal{T})$$

Then the initial position of the transformation having S_2 as control points is obtained by:

$$\mathbf{\Theta}_{2,init} = \{\mathcal{L}_1(\mathbf{\Theta}_{1,min}, P_i), P_i \in S_2\}$$

There is no previous study of the appropriate position of the control points. In practice control points are placed manually following these two basic rules, taking into account the application to shape registration:
(i) TPS being an interpolation scheme, the area of influence of a control point is limited to a restricted area in its neighborhood. In order to have better local control on the shape, every control points will be located on the shape boundary.
(ii) Control points are spaced evenly on the shape boundary, however when increasing the number of control, additional points will be located at the positions where the registration error is locally maximum.

Regarding optimization, we will use approaches which require fast computation of the transformation and its first order derivatives. Under particular conditions, all non-linear terms can be pre-computed so that the cost of a transformation is linear in the number of control points:

- A fixed set of control points is sufficient to determine the matrix L, its inverse may be pre-

computed.

- A fixed set of source points to be transformed (belongs to the source shape domain) is sufficient to determine the vector B. The product $L^{-1}.[B, \mathbf{x}', 1]$ may be pre-computed.

As a consequence, the computation of the deformation of a set of points under variations of the transformation parameter $\mathbf{\Theta}$ is performed in linear time.

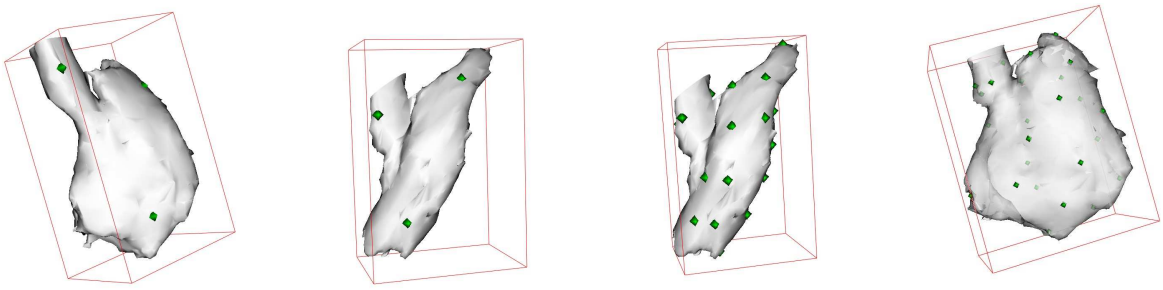


Fig. 2.7: Example of TPS Transformation of the right ventricle. From left to right: initial model with 4 control points; 4-points 3D TPS transform (equivalent to an affine transformation); Refinement of the transformation with 25 control points; TPS transform with 25 control points

Extracting Similarity transform from Thin Plate Spline

TPS is expressed as the sum of an affine transformation and local deformation under the constraint that local deformation minimize the smoothness energy in equation 2.9. The transformation being linear to vector of parameters, applying an affine transformation to a transformed point $\mathbf{x}'$ is equivalent to applying the affine transformation to the control points $\mathbf{P}'$ or to the affine and weight coefficients of the TPS:

$$\tilde{A}\mathbf{x}' + \tilde{T} = \tilde{A}A.\mathbf{x} + \tilde{A}(T + \tilde{T}) + \sum_{i=1}^{N} \tilde{A}\mathbf{w}_i G(\|P_i - \mathbf{x}\|), \tag{2.12}$$

Choosing $\tilde{A} = A^{-1}$ and $\tilde{T} = -T$ can be used to remove the affine transform from the TPS and create a model invariant under pose parameters. Also one can use this affine transform to estimate the 'closest' similarity transform so as to remove the pose parameters from the deformations. To address this issue, we have minimized the Froebenius norm $\|A - sR\|_F$ with respect to a rotation R matrix and a scaling factor s. This minimization problem admits a closed form solution [72].

Consider the SVD decomposition of the affine matrix $A = V.D.W^T$, the closest similarity is:

$$S = s.R \qquad \text{with} \qquad R = V.WT \qquad \text{and} \qquad s = trace(D)/3 \tag{2.13}$$

We consequently use $\tilde{A} = S$ and $\tilde{T} = -T$ to remove the similarity from the TPS transform. In order to obtain an efficient approximation of the similarity transform, it is is necessary to locate the center of rotation at the center point of the initial shape.

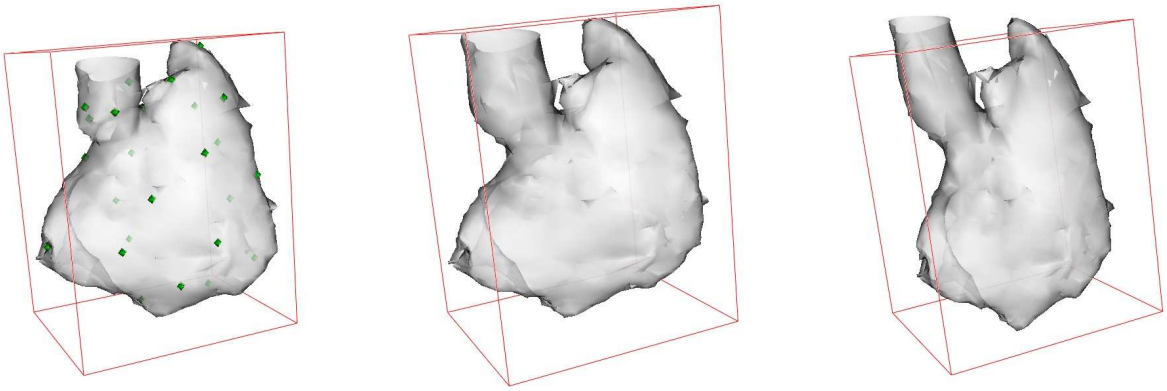


Fig. 2.8: TPS transformation with 25 control points (Right) and its best affine approximation (middle) and best similarity transform approximation (right)

2.4 Optimization

Optimization defines the process used to perform registration. In the present case, the retrieval of the best parametric transformation relies on the choice of a similarity measure between shapes and a minimization scheme. It can be directly related to the representation technique and different approaches are adopted. In most cases it can be independent of the considered transformation. Existing methods often decompose the process in two steps. First correspondences between the source and the target are established and then the transformation parameters for known correspondences are recovered.

Iterative closest point is a widely used technique to register shapes represented by sets of points along with a parametric transformation. Consider two sets of points $\mathcal{P} = \{\mathbf{p}_i\}$ from the source shape boundary and $\mathcal{Q} = \{\mathbf{q}_i\}$ from the target shape boundary. Consider a certain parametric transformation $T(\mathbf{\Theta}, .)$, the problem is to find the transformation parameters $\mathbf{\Theta}$, and associate a set of correspondences $\mathcal{C} \subset \mathcal{P} \times \mathcal{Q}$, that minimize an appropriate error distance. A general objective

function can be written as:

$$E(\mathbf{\Theta}, \mathcal{C}) = \sum_{(\mathbf{p}_i, \mathbf{q}_j) \in \mathcal{C}} \rho(d(\mathcal{L}(\mathbf{\Theta}; \mathbf{p}_i), \mathbf{q}_j)), \tag{2.14}$$

where $d(\mathcal{L}(\mathbf{\Theta}; \mathbf{p}_i), \mathbf{q}_j)$ is a distance between a transformed source point and its corresponding target point and ρ is a loss function. Euclidean distance is a common choice for d [13].

A common evolution is assuming a point to curve distance following two different approaches. On the one hand it is assuming that the set of the model points $\mathcal{P}$ is provided with normal information $(\mathbf{p}_i \leftrightarrow \eta_i)$, this was developped in [30]. On the other hand it may be assumed that the set of data points $\mathcal{Q}$ is sampled from a smooth interface and provided with normal information $(\mathbf{q}_j \leftrightarrow \eta_j)$ [173]:

$$d(M(\mathbf{\Theta}; \mathbf{p}_i), \mathbf{q}_j) = |d(M(\mathbf{\Theta}; \mathbf{p}_i) - \mathbf{q}_j)^T . \eta_j|. \tag{2.15}$$

A usual choice for ρ is leading to the mean square error ($\rho(u) = u^2$), whose minimization has a closed form solution. However more robust approach were used to reject or at least weaken the influence of mismatched points using M-estimators [85]. Standard ICP is considering a certain class of transformation and distance functions, and iteratively minimizes the above functional with respect to the parameters for a fixed set of correspondences, and estimating point correspondence for a given transformation.

More evolved techniques address the issues of robustness using more complex transformations to improve the quality of the correspondences. Softassign techniques introduced in [146] allow to relax the correspondence problem by introducing a double stochastic matrix M. The above distance is then computed for all couples of points in $\mathcal{P} \times \mathcal{Q}$:

$$E(\mathbf{\Theta}) = \sum_{(\mathbf{p}_i, \mathbf{q}_j) \in \mathcal{P} \times \mathcal{Q}} M_{i,j} \rho(d(T(\mathbf{\Theta}; \mathbf{p}_i), \mathbf{q}_j)) + F(M)/T, \tag{2.16}$$

where $M_{i,j}$ correspond to the weight associated to the correspondence of the points $(\mathbf{p}_i, \mathbf{q}_j)$ and summing to 1 along the both index i and j ($\sum_i M_{i,j} = 1$ and $\sum_i M_{i,j} = 1$). F is an entropy term ($F(M) = \sum_{i,j} M_{i,j} log(M_{i,j})$) which preserves positive coefficients for M and enforces the matrix to contains uniquely 0 and 1 (minimal entropy) in a particular framework inspired from simulated annealing when the "temperature" T goes to 0. In the case of a quadratic distance, this approach sums up to the registration of each point on the source shape $\mathbf{p}_i$ to a weighted sum of the target points ($\sum \alpha_j \mathbf{q}_j$) where the coefficients α_i are a function of the distance $d(M(\mathbf{\Theta}; \mathbf{p}_i) - \mathbf{q}_j)$ and the temperature $Temp$. This is of particular interest if the points $\mathbf{q}_j$ are extracted from a continuous interface as interpolated points will be located approximately on the interface. This

approach was also combined with a thin plate spline transformation in [34] where the source data points $\mathbf{p}_i$ are chosen as control points, and the coarse to fine approach in the degree of freedom of the transformation is uniquely provided by the regularization term (section 2.3.2). The main limitation of these methods refers to the decomposition of the process in two steps and the explicit need of recovering correspondences. Following a very different approach, shape registration based on distance transform was introduced by Rousson and Paragios in [134], having in mind that this approach would consists of registering different iso-level of both source and target shape. Therefore the energy can be formulated with the general functional:

$$E(\mathbf{\Theta}, \mathcal{S}, \mathcal{T}) = E_{data}(\phi_{\mathcal{S}}(.), \phi_{\mathcal{T}} \circ \mathcal{L}(\mathbf{\Theta}, .)), \tag{2.17}$$

where E_{data} refers to a similarity measure between distance functions, $\phi_{\mathcal{S}}$ and $\phi_{\mathcal{T}}$ the distance functions to respectively $\mathcal{S}$ and $\mathcal{T}$. This approach is reverse: instead of minimizing a criterion between the transformed source $(\mathcal{L}(\mathbf{\Theta}, \mathcal{S}))$ and the target $\mathcal{T}$, we are considering the distance transformed $\phi_{\mathcal{T}} \circ \mathcal{L}(\mathbf{\Theta}, .)$, that implicitly use the inverse of the transformation:

$$\phi_{\mathcal{T}} \circ \mathcal{L}(\mathbf{\Theta}, \mathbf{x}) = 0 \quad \Leftrightarrow \quad \mathcal{L}(\mathbf{\Theta}, \mathbf{x}) \in \delta\mathcal{T} \tag{2.18}$$

$$\Leftrightarrow \quad \mathbf{x} \in \mathcal{L}^{-1}(\mathbf{\Theta}, \mathcal{T}) \tag{2.19}$$

Therefore this approach is comparing the inverse transform of $\mathcal{T}$ with the source shape $\mathcal{S}$. The initial approach in [134, 135] makes use of L^2 distance on the distance functions, and optimizes with respect to a similarity transform and an unconstrained non-rigid deformation. Using a general loss function ρ on $\mathbb{R}$, the energy is generalized as:

$$E(\mathbf{\Theta}, \mathcal{S}, \mathcal{T}) = \int_{\Omega} \rho(\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x})))d\mathbf{x}, \tag{2.20}$$

bringing two additional modifications:

- similarity transform : the energy is modified to preserve the distance constraint on the transformed target distance function $\phi_{\mathcal{T}} \circ T$ following the approach demonstrated in section 2.2.4.

$$E(\mathbf{\Theta}, \mathcal{S}, \mathcal{T}) = \int_{\Omega} \rho(\phi_{\mathcal{S}}(\mathbf{x}) - (\phi_{\mathcal{T}}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x}))/|J_T(\mathbf{\Theta}, \mathbf{x})|^{1/d}))d\mathbf{x}, \tag{2.21}$$

  where $|J_T(\mathbf{\Theta}, \mathbf{x})|$ is the determinant of the Jacobian matrix of the transformation. $|J_T(\mathbf{\Theta}, \mathbf{x})|^{1/d}$ corresponding to the scaling factor of the similarity.

- local transform: the above registration energy is combined with an elastic regularization term

E_{smooth} as presented in section 2.3.2.

This approach was then extended to local parametric transformation using FFD in [83], and independently the similarity measure between distance transform changed to mutual-information, following the approach introduced by Viola and Wells in [203].

The next section will show how these two different approaches are related before introducing our registration framework.

2.4.1 Optimization Strategies

Consider a point $\mathbf{x}$ from the source reference frames, its distance value is $\phi_{\mathcal{S}}(\mathbf{x})$ and it is transformed through the registering transformation to the target domain as $\mathbf{x}' = \mathcal{L}(\mathbf{\Theta}, \mathbf{x})$. It can be shown that the minimization of $|\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathbf{x}')|$ with respect to $\mathbf{x}'$ is equivalent to the minimization of the distance between $\mathbf{x}'$ and its projection on the level set of $\phi_{\mathcal{T}}$ with value $\phi_{\mathcal{S}}(\mathbf{x})$.

However there are a few restrictions. $\phi_{\mathcal{T}}$ being produced from a smooth manifold, it is differentiable almost everywhere. Considering signed distance functions, $\nabla\phi_{\mathcal{T}}$ is not defined on the internal and external skeleton of the shape $\mathcal{T}$ (Skeleton is defined as the loci of the centers of all the bitangent circles to the boundary of the shape [14]). Assume that $\phi_{\mathcal{T}}$ is differentiable at $\mathbf{x}'$, consider $P_{\delta\mathcal{T}}(\mathbf{x}')$ its projection on the boundary of the target shape. It is interesting to notice that for any point located along the line $(\mathbf{x}', P_{\delta\mathcal{T}}(\mathbf{x}'))$ between the internal and external skeleton, its projection on the boundary of $\mathcal{T}$ remains identical. Formally, this can be written:

$$P_{\delta\mathcal{T}}(\mathbf{x}' + \alpha.\mathbf{u}(\mathbf{x}')) = P_{\delta\mathcal{T}}(\mathbf{x}') \qquad \text{with} \qquad \mathbf{u}(\mathbf{x}') = \mathbf{x}' - P_{\delta\mathcal{T}}(\mathbf{x}') \tag{2.22}$$

With the above definition of $\mathbf{u}$, also called Vector distance function [73], this proves the relation:

$$\begin{aligned} \mathbf{u}(\mathbf{x}' + \alpha.\mathbf{u}(\mathbf{x}')) &= \mathbf{x}' + \alpha.\mathbf{u}(\mathbf{x}') - P_{\delta\mathcal{T}}(\mathbf{x}') \\ &= (1+\alpha)\mathbf{u}(\mathbf{x}') \end{aligned}$$

As we also have the relation between distance and vector distance:

$$\mathbf{u}(\mathbf{x}') = \phi_{\mathcal{T}}(\mathbf{x}').\nabla\phi_{\mathcal{T}}(\mathbf{x}'), \tag{2.23}$$

These two relations shows that the gradient of the distance function remains constant along the line $(\mathbf{x}', P_{\delta\mathcal{T}}(\mathbf{x}'))$. This relation demonstrates the redundancy of the distance map representation: any

level set can be used to recover (at least partially) the original shape. Also, the projection of $\mathbf{x}'$ on any other iso-level of $\phi_{\mathcal{T}}$ with value located between the inner and outer skeletton is also located on this line. Consequently, and following these restrictions, the distance between level sets can be expressed as follows:

$$\begin{aligned} |\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathbf{x}')| &= \|\mathbf{x}' - (\mathbf{x}' + (\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathbf{x}')).\nabla\phi_{\mathcal{T}}(\mathbf{x}'))\| && (2.24)\\ &= \|\mathbf{x}' - (P_{\delta\mathcal{T}}(\mathbf{x}') + \phi_{\mathcal{S}}(\mathbf{x}).\nabla\phi_{\mathcal{T}}(\mathbf{x}'))\| && (2.25)\\ &= \|\mathbf{x}' - (P_{\{\mathbf{y}|\phi_{\mathcal{T}}(\mathbf{y})=\phi_{\mathcal{S}}(\mathbf{x})\}}(\mathbf{x}'))\|, && (2.26) \end{aligned}$$

where $P_{\{\mathbf{y}|\phi_{\mathcal{T}}(\mathbf{y})=\phi_{\mathcal{S}}(\mathbf{x})\}}$ represents the projection on the iso-level of $\phi_{\mathcal{T}}$ with value $\phi_{\mathcal{S}}(\mathbf{x})$. Let $\mathbf{p}_{\mathbf{x}'} = P_{\{\mathbf{y}|\phi_{\mathcal{T}}(\mathbf{y})=\phi_{\mathcal{S}}(\mathbf{x})\}}(\mathbf{x}')$ be this projection, it is also interesting to differentiate the above quantity for small variations of $\mathbf{x}'$:

$$\begin{aligned} \phi_{\mathcal{T}}(\mathbf{x}' + \delta\mathbf{x}') - \phi_{\mathcal{S}}(\mathbf{x}) &= \phi_{\mathcal{T}}(\mathbf{x}') + \nabla\phi_{\mathcal{T}}(\mathbf{x}').\delta\mathbf{x}' - \phi_{\mathcal{S}}(\mathbf{x}) && (2.27)\\ &= \phi_{\mathcal{T}}(\mathbf{x}') - \phi_{\mathcal{T}}(\mathbf{p}_{\mathbf{x}'}) + \nabla\phi_{\mathcal{T}}(\mathbf{p}_{\mathbf{x}'}).\delta\mathbf{x}' && (2.28) \end{aligned}$$

This last equation shows that up to the first order, the quantity $\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathbf{x}')$ refers to a point to line distance. With this approach, registration using distance functions can be directly compared to ICP methods where closest points are considered on different iso-levels of the distance function. This can be performed by optimizing the similarity measure with respect to the parameter vector, considering point projection on the different iso-level line at the previous step:

$$\boldsymbol{\Theta}_{i+1} = \arg\min_{\boldsymbol{\Theta}} \int_{\Omega} \rho\left(\phi_{\mathcal{T}}(\mathcal{L}(\boldsymbol{\Theta}_i, \mathbf{x})) - \phi_{\mathcal{S}}(\mathbf{x}) + \left[\mathcal{L}(\boldsymbol{\Theta}, \mathbf{x}) - \mathcal{L}(\boldsymbol{\Theta}_i, \mathbf{x})\right].\nabla\phi_{\mathcal{T}}(\mathcal{L}(\boldsymbol{\Theta}_i, \mathbf{x}))\right) d\mathbf{x}. \tag{2.29}$$

This admits a closed form solution in the case of a quadratic cost function ρ

Optimization algorithms

The approach used in most of the shape registration work is based on the calculus of the variations of the energy and optimized using gradient descent:

$$\frac{d\boldsymbol{\Theta}}{dt} = \frac{dE}{d\boldsymbol{\Theta}}(\boldsymbol{\Theta}) = \int_{\Omega} \rho'(\phi_{\mathcal{T}}(\mathcal{L}(\boldsymbol{\Theta}, \mathbf{x})) - \phi_{\mathcal{S}}(\mathbf{x})) < \frac{d\mathcal{L}}{d\boldsymbol{\Theta}}(\boldsymbol{\Theta}, \mathbf{x}) | \nabla\phi_{\mathcal{T}}(\mathcal{L}(\boldsymbol{\Theta}, \mathbf{x})) > d\mathbf{x} \tag{2.30}$$

$$\boldsymbol{\Theta}_{t+dt} = \boldsymbol{\Theta}_t + \frac{dE}{d\boldsymbol{\Theta}}(\boldsymbol{\Theta}).dt \tag{2.31}$$

This approach is supposed to lead to a monotonically decreasing energy for an appropriate step size dt. However the time step dt that preserve a decreasing sequence of energy value may vary according to the position. We use a simple method which enforces a strictly decreasing energy by dividing the time step by a factor μ_2 ($\mu_2 > 1$) for every ascending step starting from the previous position. In order to speed up the descent, the time step is multiplied by a factor μ_1 at every successfully descending step. In our application we have used $\mu_1 = 1.2$ and $\mu_2 = 1.5$.

A more evolved technique is based on second order development of the energy with respect to the parameter $\mathbf{\Theta}$ and therefore makes use of its Hessian matrix. This is known as *Marquardt-Levenberg* method:

$$\mathbf{\Theta}_{i+1} = \mathbf{\Theta}_i - (\mathcal{H}_E(\mathbf{\Theta}) + \lambda I)^{-1} \nabla E(\mathbf{\Theta}), \tag{2.32}$$

where $\mathcal{H}_E$ refers to the Hessian of E, and λ is a scalar parameter that modifies the algorithm from pure Newton method ($\lambda \approx 0$) or simple gradient descent ($\lambda >> 1$). Following the previous approach, λ is divided by μ_1 after a successful descent step, and multiplied by μ_2 after a failing step. Within our approach the Hessian used in the above expression is actually a first order approximation computed with the approach presented in section 2.5

2.4.2 Narrow Bands and Distance Reinitialization

Such a variational method is quite sensitive to the initial conditions and could converge to a local minimum. One also notices that the proposed form of energy cannot lead to exact results: it was shown in section 2.2.4 that the distance transformed can be preserved only in the case of similarity transformations. For more general transformations, the gradient is distorted with the Jacobian matrix of the transformation and its norm cannot be kept equal to one (eikonal constraint). Considering the classic relation for any $\mathcal{C}^1$ scalar function defined on Ω and any $\mathcal{C}^1$ parametric contour $C(t)$ with $C(0) = \mathbf{x}_0$ and $C(1) = \mathbf{x}_1$:

$$\phi_{\mathcal{T}}(\mathbf{x}_1) = \phi_{\mathcal{T}}(\mathbf{x}_0) + \int_0^1 < \frac{dC}{dt}(t) | \nabla \phi(C(t)) > dt. \tag{2.33}$$

With $\phi(\mathbf{x}) = \phi_{\mathcal{T}}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x}))$, and C the segment joining any point $\mathbf{x}$ to its projection on the transformed shape ($\mathbf{y} = P_{\delta\mathcal{L}^{-1}(\mathbf{\Theta},\mathcal{T})}(\mathbf{x})$). $\mathbf{y}$ satisfies $\phi(\mathbf{y}) = 0$ and shows that deformations of the gradient are integrated along the segment and the further $\mathbf{x}$ is located from the zero iso-level, the more the distance will be distorted. Therefore in [135], the idea of considering a narrow band

surrounding the source shape was proposed to make the registration process more robust.

$$E_\alpha(\mathcal{L}(\mathbf{\Theta})) = \int_\Omega \chi_\alpha(\phi_\mathcal{S}(\mathbf{x}))\rho\left(\phi_\mathcal{S}(\mathbf{x}) - \phi_\mathcal{T}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x}))\right) d\mathbf{x} \tag{2.34}$$

with χ_α being the characteristic/indicator function:

$$\chi_\alpha(x) = \begin{cases} 1/(2\alpha) & \text{if } x \in [-\alpha, \alpha] \\ 0 & \text{else} \end{cases} \tag{2.35}$$

The selection of the parameter α is crucial since to some extent it refers to the scale of the shapes to be registered. In addition, it is natural when converging to the optimal solution that α tends to 0. Therefore, we assume a finite number of decreasing values of scales associated to narrow band radii $\{\alpha_0 > ... > \alpha_t > ... > \alpha_n \approx 0\}$ that is equivalent to a scale decomposition of the process.

The same kind of attention should be paid to the choice of transformation: if Θ is too large, there is a high risk of converging to a local minimum. So, while updating the scale value, we also progressively increase the complexity of the transformation and therefore the number of parameters in the vector $\mathbf{\Theta}$.

In order to account for the non invariance of the distance transform under basic transformations, the distance to the transformed source shape is recomputed during the minimization process when updating the parameter α_t.

Formally, let $\mathbf{\Theta}_{t-1}$ be the parameters defining the transformation $\mathcal{L}_{t-1}(.) = \mathcal{L}(\mathbf{\Theta}_{t-1}, .)$ for which the energy was minimum at scale $t-1$. Also let $\mathcal{S}^{t-1} = \mathcal{L}_{t-1}(\mathcal{S})$, that is the source shape transformed with the current transformation . The registration between shapes is then equivalent to iteratively minimizing:

$$E_{\alpha_t}(\mathcal{L}(\mathbf{\Theta})) = \int_\Omega \chi_{\alpha_t}(\phi_\mathcal{S}(\mathbf{x}))\rho(\phi_{\mathcal{S}^{t-1}}(\mathcal{L}_{t-1}(\mathbf{x})) - \phi_\mathcal{T}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x})))d\mathbf{x} \tag{2.36}$$

where $\phi_{\mathcal{S}^{t-1}}()$ refers to the distance transform of the shape $\mathcal{L}_{t-1}(\mathcal{S})$. Within such a formulation the integration domain is always related to the initial source shape and does not depend on the number of iteration or the parameter α_t. Moreover when α_t tends to 0, $E_{\alpha_t}(\mathcal{L}(\mathbf{\Theta}))$ is equivalent to the point based registration on shape boundary:

$$E_{\alpha_\infty}(\mathcal{L}(\mathbf{\Theta})) = E_0(\mathcal{L}(\mathbf{\Theta})) = \int_{\delta\mathcal{S}} \rho(\phi_\mathcal{T}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x})))d\mathbf{x}. \tag{2.37}$$

2.4.3 Similarity Measures

Most of the work on shape registration using distance transforms use the L^2 norm, that is a quadratic cost function ρ. This choice is sensitive to outliers and therefore enforces the importance of the narrow band: a large size will produce larger values of the distance and these points potentially erroneous will have a greater role in the optimization process. Less sensitive cost functions can be used, like absolute value. Also ρ could be chosen equal to an M-estimator like Beaton-Turkey so that large values of the difference $\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x}))$ will not have any influence on the optimization.

A different approach was used to address the problem of distance deformation upon a transformation. In [84] the similarity measure between $\phi_{\mathcal{S}}$ and $\phi_{\mathcal{T}} \circ \mathcal{L}$ makes use of mutual information. This method considers that the distance values of $\phi_{\mathcal{S}}$ and $\phi_{\mathcal{T}} \circ \mathcal{L}$ at every pixels are independent samples from 2 unknown random variables and the transformation $\mathcal{L}$ as the one that minimizes the independence of the random variables. There exist various ways to measure independence of random variables, mutual Information being one of them, it was introduced in the case of multi-modal image registration in [203]:

$$E(\mathbf{\Theta}) = H(\phi_{\mathcal{S}}) + H(\phi_{\mathcal{T}} \circ \mathcal{L}) - H(\phi_{\mathcal{S}}, \phi_{\mathcal{T}} \circ \mathcal{L}(\mathbf{\Theta}, .)), \tag{2.38}$$

$$H(\phi_{\mathcal{S}}) = \int p_{\phi_{\mathcal{S}}}(t) \log(p_{\phi_{\mathcal{S}}}(t)) dt \quad \text{where} \quad p_{\phi_{\mathcal{S}}}(t)dt = \int_{\Omega} \chi_{[t,t+dt]} \phi_{\mathcal{S}}(\mathbf{x}) d\mathbf{x}/|\Omega|$$

$$H(\phi_{\mathcal{T}} \circ \mathcal{L}) = \int p_{\phi_{\mathcal{T}} \circ \mathcal{L}}(t) \log(p_{\phi_{\mathcal{T}} \circ \mathcal{L}}(t)) dt \quad \text{where} \quad p_{\mathcal{T} \circ \mathcal{L}}(t)dt = \int_{\Omega} \chi_{[t,t+dt]} \phi_{\mathcal{T}}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x})) d\mathbf{x}/|\Omega|$$

$$H(\phi_{\mathcal{S}}, \phi_{\mathcal{T}} \circ \mathcal{L}(\mathbf{\Theta}, .)) = \int p_{\phi_{\mathcal{S}}, \phi_{\mathcal{T}} \circ \mathcal{L}}(u, t) \log(p_{\phi_{\mathcal{S}}, \phi_{\mathcal{T}} \circ \mathcal{L}}(u, t)) du dt, \tag{2.39}$$

where the last expression represents the entropy of the joint probability associated with $\phi_{\mathcal{S}}$ and $\phi_{\mathcal{T}} \circ \mathcal{L}$. In practice integration is not performed in the continuous domain but the output values of the distance functions are sorted into histograms so that entropies are computed on discrete random variables.

This approach has shown better results in the case of large deformation with a few degrees of freedom for the transformation (affine, similarity). These particular cases lead to an inexact match of the shape boundary and their associated distance transforms. While Standard methods (SSD, SAD, M-estimator) will try to match every level-set of the distance transforms, one notices that mutual information criterion aligns images so that a functional relation exists between the values

of $\phi_{\mathcal{S}}$ and $\phi_{\mathcal{T}} \circ \mathcal{L}$. Therefore the scaling correction factor applied to preserve a distance function in the case of a similarity is useless when using this criterion. In the case of local deformations, the interest of mutual information is weakened when compared to classic approaches for 2 reasons. (i) The warping between shape boundary being of better quality, the match between distance proved sufficient to consider a measure on their difference. (ii) Local deformation is associated to a smaller narrow band in the optimization process, therefore less prone to distance deformation.

One of the main criticism of geometric registration in medical imaging is that such correspondences do not always refer to meaningful anatomical ones. The use of landmarks, automatically or manually extracted, is a prominent step to introduce anatomical consistency.

2.4.4 Landmark Registration

Some shapes are given along with landmarks positions. We use this information to enforce the correspondences between the shapes to register and amend the registration energy. Landmarks positions being exact, we have used a simple additional quadratic term in the energies.

$$E_{landmarks}(\mathbf{\Theta}) = \sum_{i=1}^{N} \|\mathcal{L}(\mathbf{\Theta}, \mathbf{p}_i) - \mathbf{q}_i\|^2, \tag{2.40}$$

where $(\mathbf{p}_1 \dots \mathbf{p}_N)$ is a set of N landmarks located on the source shape, $(\mathbf{q}_1 \dots \mathbf{q}_N)$ the set of N corresponding landmarks on the target frame.

2.4.5 Implementation Remarks

The global energy is composed of 3 different terms:

$$E(\mathbf{\Theta}) = \frac{1}{a+b+c}(aE_{\alpha,shape}(\mathbf{\Theta}) + bE_{smooth}(\mathbf{\Theta}) + cE_{landmarks}(\mathbf{\Theta})), \tag{2.41}$$

where $E_{\alpha,shape}$ refers to the distance based registration energy, and (a, b, c) are mixing scalar positive parameters. $E_{\alpha,shape}$ contains the information on the narrow band (with size α) used to register distances transforms (section 2.4.2). E_{smooth} may only be used in the case of local registration (section 2.3.2). $E_{landmarks}$ is mostly used in the first steps of the registration to help recover large deformations. The value of c decreases when considering more local deformations and smaller narrow bands.

Discretization of the integrals involved in $E_{\alpha,shape}$ is performed on a regular grid with variable spacing lying on the source domain. Points of the grid located within the narrow band are considered. The computation of energy and gradients being linear in the number of points, the process initially uses a coarse grid associated with a large narrow band, then refines the grid as the size of the narrow band is decreasing so that the global number of points involved remains constant.

It is important to notice that registration carried on distance transforms contains certain redundancy. This variable sparseness of the discretization is also of interest to address this issue. If considering a large band, the registration problem is formulated on a large set of iso-contours having coherent normal directions related to the curvature of the target shape (section 2.4.1). Upon discretization, using numerous points located along the same line orthogonal to the target boundary (namely a line joining $\mathbf{x}'$ and $P_{\delta\mathcal{T}}(\mathbf{x}')$) which implicitly addresses the registration of the same 0-level position of the source shape boundary and is therefore redundant. Sparseness addresses this limitation by spreading the point projection all along the shape boundary.

Regarding optimization, the grid being fixed, it has been shown in section 2.3.2 that all non-linear computation involved in the local transformation can be pre-computed so that registration complexity is also linear in the number of control points.

Setting the mixing sclars values (a, b, c) mostly depends on the application. A global approach consists in favoring a large smoothing coefficients in the first steps, therefore performing rough segmentation (with affine deformation being preferred to local deformations). Then decreasing the importance of this value leads to more local and softer deformations. Meanwhile landmarks information (if available) plays a prominent role in the first steps of the process to align properly small anatomical structures. Along the process its influence is decreased so that the shape term alone has significant role in the final refinement steps.

2.4.6 Experimental results

The registration framework has been experimented on various examples of 2D and 3D data with different possible application. Large scale experimentation were performed having in mind the future construction of a probabilistic model of the deformations. We have first considered digits registration which shows very large local deformations (see appendix C). However regarding medical applications it was experimented on sagittal view of the corpus callosum, with training sample manually extracted from MR-images. Minimization of the registration energy is performed using gradient descent, where successively refine the size of the narrow band from .3 to .05 times the size of the shape, and simultaneously increase the complexity of the diffeomorphism (from an affine

transformation to an FFD with a regular $[12 \times 7]$ lattice) as shown on figure 2.9.

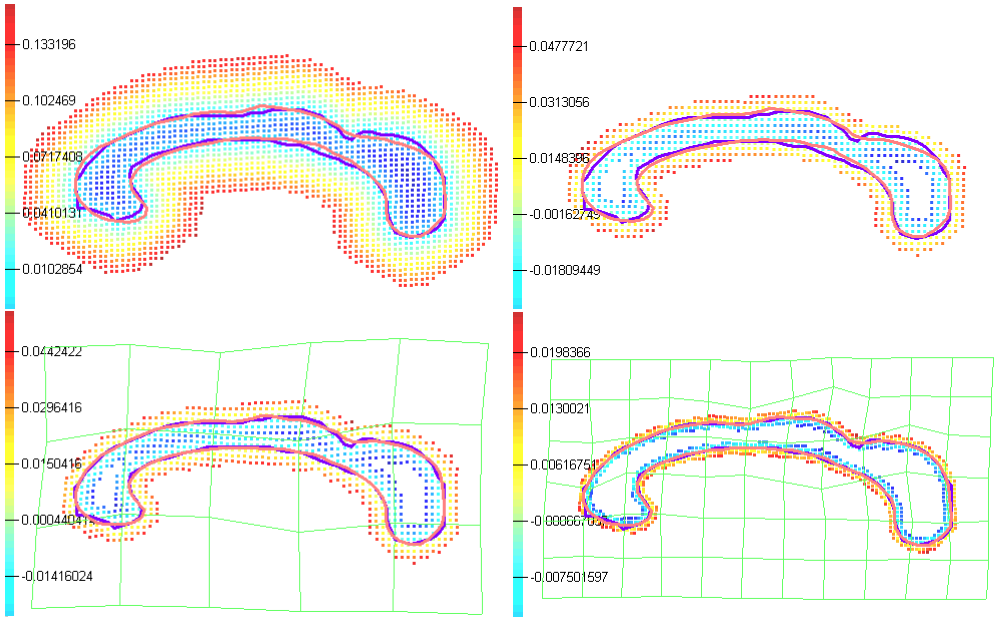


Fig. 2.9: Full Registration process on Corpus Callosum with FFD. Registration using similarity transform (top right), registration using affine transform (top left), registration using 6×4 FFD lattice, registration using 12×7 FFD lattice

Our registration framework was also experimented on 3D shapes of cardiac left ventricle. We have used a set of segmented CT and MRI volumes, including both healthy and pathological hearts. This set of shapes contains 25 examples extracted from CT volumes with the semi-automatic method presented in [74] and 18 examples manually segmented from MR-images.

When data is acquired with MRI, the scan domain is restricted to the most important parts of the ventricle. Consequently, apex, mitral and aortic valves are not present and the segmented ventricle represent a truncated shape (Fig. 2.10). The algorithm has been slightly modified to register this type of truncated data, so that registered points located outside of the target volume do not contribute to the energy. The cost function being quadratic, registration is performed considering successively similarity transform, affine and TPS transform with 60 and then 90 control points. Simultaneously, the narrow band considered is decreasing from from .3 to .05 times the size of the shape. Initialization is performed by aligning the center of mass of the source and target shape, and aligning the direction of the lung, on the images. Such initialization process approximately aligns the aortic valves of both ventricles so that the registration process robustly converges to the desired results (Fig. 2.11).

Our registration framework was experimented on various type of data as shown on figure 2.17:

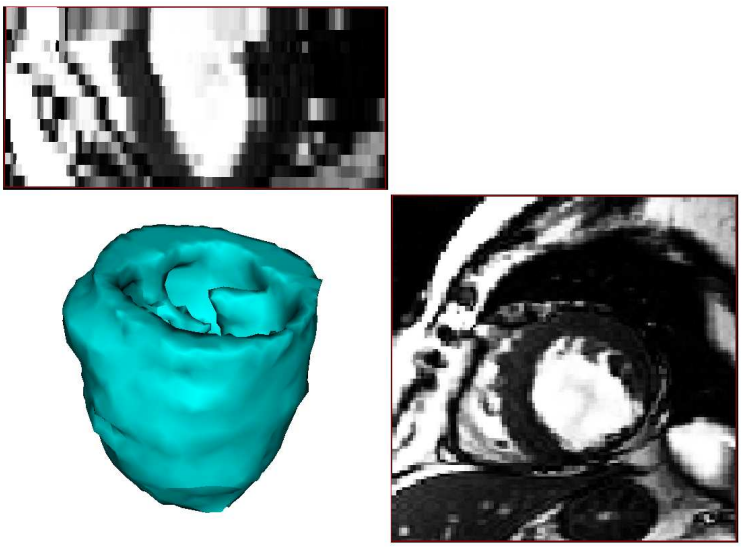


Fig. 2.10: Representation of the manually segmented left ventricle from MRI data

- 5 class segmentations of the heart (left/right ventricles, left/right atria and aorta) with 3D free form deformation. It reaches the maximum accuracy allowed with the size of the grid.
- brain cortex segmentation. In such case segmentation align large structure but fails to register the numerous foldings correctly.
- 3D faces: Synthetic meshes, data extracted with a laser scanner (both clouds of points and reconstructed meshes), face extracted from a binocular video system [181]
- Full body meshes extracted from laser scanner.

The proposed framework would be capable to produce dense correspondences between any two surfaces through the minimization of the proposed objective function. However the objective function is highly non convex and therefore convergence to a local minimum is possible. One can overcome this limitation either through the use of more advanced optimizers, or through the estimation of the local uncertainty of the obtained solution.

2.5 Estimation of Registration Uncertainties

Several attempts to build statistical models on noisy sets of data in order to infer the properties of a certain model have been proposed in the literature. There is a persistent relation in the literature between noise and uncertainty. Distinction is more semantic than mathematical: One could simply

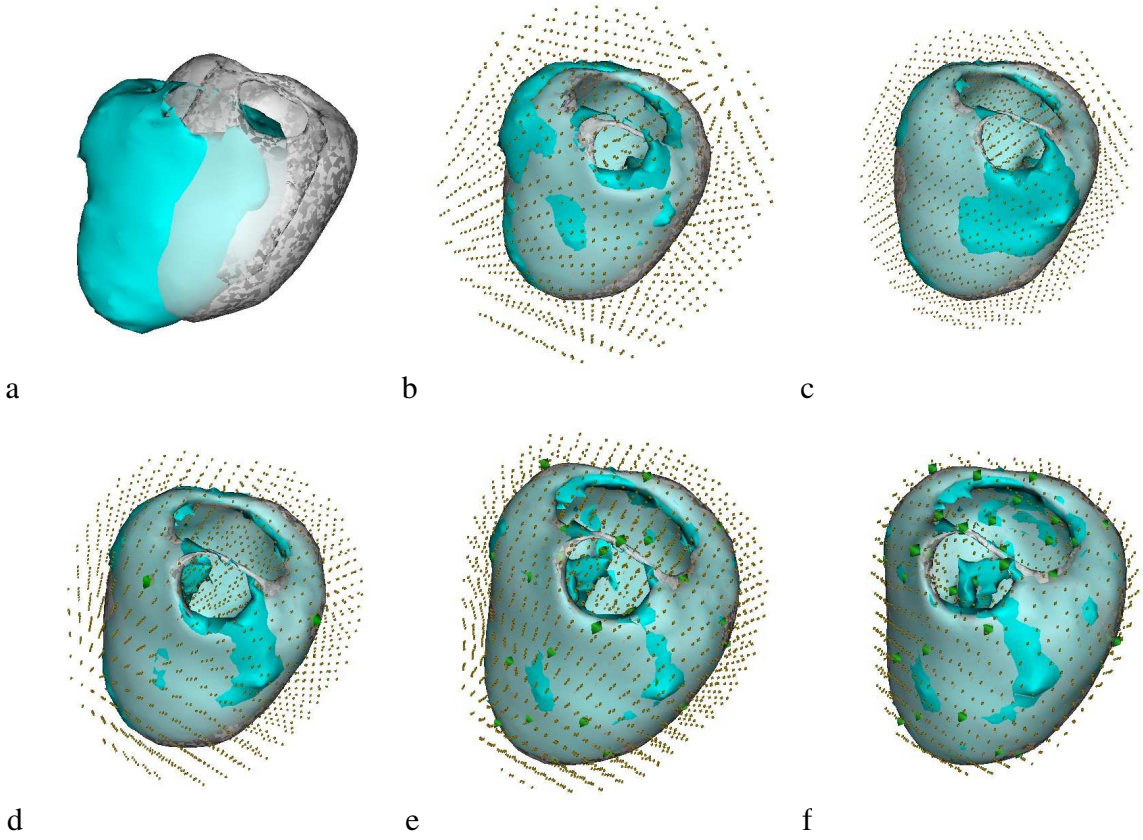


Fig. 2.11: Full Registration process on cardiac left ventricle. (a) Initial pose. (b) After automatic initialization process. (c) Similarity transform. (d) 4 control points TPS transformation. (e) 60 control points TPS-transform. (f) 90 control points TPS-transform.

state that noise is likely to be removed as uncertainty is likely to be estimated, both cases refer to an unknown random variable. In the present case, registration of a geometric shape model is performed, aligned with a noisy complex target shape. Registration may not be exact and uncertainty on the retrieved transformation will provide additional information. This approach is related to different work finding the source of uncertainty in landmarks points estimation and uses it to infer different global parameters.

2.5.1 State of the Art

In [91], Kanatani reports various techniques to extract feature points in images along with uncertainties due to the inherent noise. These information being used for geometric fitting of structures. In real world, any measurement is subject to uncertainty. Considering a set of measurement or samples is the starting point to statistical estimation to infer parameters of the measured variables.

Origin of uncertainty can be divided in two categories. It can be either external, which means it originates from the quality of the measuring devices (it depends on the noise level or image quality) It can be internal in which case it originates from the physical phenomenon itself and the only way to reduce it requires a repetition of the measurement procedure and doing statistical inference. In the context of [91] where feature points need to be extracted from images, uncertainty is internal because it directly originates from the algorithm used to determine feature points. Consequently there is no other way to decrease uncertainty than multiplying the experiences with different feature points extraction algorithms. Then discussing the fact that there is only one experiment available, their approach builds the model to infer uncertainty on such a limit case. Regarding the registration process, one can imagine several tests of the method using different initial conditions, optimization techniques and similarity measures. Then uncertainty can be inferred according to the variance of the obtained solutions. Such an approach is mathematically well founded but not-usable in practice.

A similar idea was used for landmarks registration with a Thin Plate Spline functional in [149]. Two sets of landmarks are given on two images either with orientation information (an additional vector is given if landmarks are located on a visible boundary of an object) or given with uncertainty information (a covariance based on gradient and structure matrix is given if landmarks is located on a noisy position). Image registration is performed by updating the registration energy with this information, however, no uncertainty is inferred on the final result.

In [139], an iterative estimation method was proposed to handle uncertainty estimates of rigid motion on sets of matched points. Their contribution is due to the fact that they no longer consider additive noise but place the noise at the center of the algorithms, it subsequently presents all mathematical basis for the propagation of noise model through explicit or implicit functions. Their work is directly related to ours as we are also projecting uncertainties along shape border to the parameters of the transformation.

In a quite different context, in [166], Simoncelli introduced uncertainties within the estimation of dense optical flow modeled with 2D Gaussian densities. This was based on a probabilistic model on the image gradient and temporal derivative. Being able to retrieve large displacements, such algorithm can also be seen as an image registration process where image at time t needs to be registered on image at time $t + 1$. In the present case shapes are considered using implicit representations, therefore uncertainty lie in the relative position of isosurfaces. Consequently the uncertainty in the registration problem can be seen as equivalent to the ”aperture problem” in optical flow estimation.

Last, but certainly not least, the closest related work can be found in [174] where an iterative

technique to determine uncertainties on point to curve registration was proposed. Within their original ICP based approach, correspondence between source and target is based on closest points and used to compute an objective function based on the parameters of a parametric transform. This functional is used to compute a covariance matrix ($\Sigma_{\mathbf{\Theta}}$) error on the parameter estimate. The dual bootstrap is an approach that uses this covariance estimate to select the most appropriate transformation model, and parallel to that select the spacial area of images where feature points are selected and considered to compute the objective function.

Recently in [168] an original approach was introduced to relax the constraint on points correspondence between source and target sets of points. Their methods is extending the softassign method of [33] (Eq. 2.16) where all correspondences between pairs of points are considered. The original softassign equation is based uniquely on the Euclidean distance between points (transformed source and target) to determine the influence of a pair with the matrix of coefficients M_{ij}:

$$E(\mathbf{\Theta}) = \sum_{(\mathbf{p}_i,\mathbf{q}_j)\in\mathcal{P}\times\mathcal{Q}} M_{i,j}\rho(d(T(\mathbf{\Theta};\mathbf{p}_i),\mathbf{q}_j)) + F(M)/T, \tag{2.42}$$

In [168], computation of weights is replaced with a more evolved estimate which also considers transformation error (estimated by $\Sigma_{\mathbf{\Theta}}$). Within their approach, the registration energy is a function of the parameter of the transformation Θ and its covariance matrix $\Sigma_{\mathbf{\Theta}}$:

$$E(\mathbf{\Theta},\Sigma_{\mathbf{\Theta}}) = \sum_{(\mathbf{p}_i,\mathbf{q}_j)\in\mathcal{P}\times\mathcal{Q}} M_{ij}(\rho_{ij})\underbrace{\rho(d(T(\mathbf{\Theta},\mathbf{p}_i),\mathbf{q}_j,\Sigma_{\mathbf{\Theta}}))}_{\rho_{ij}}, \tag{2.43}$$

where $\rho_{ij} = d(T(\mathbf{\Theta},\mathbf{p}_i),\mathbf{q}_j,\Sigma_{\mathbf{\Theta}})$ is a Mahalanobis error distance, which accounts for $\Sigma_{\mathbf{\Theta}}$ to compute local error covariance associated to the couple $(\mathbf{p}_i,\mathbf{q}_j)$. The matrix of weights on correspondences $M_{i,j}$ is also a function of the Mahalanobis distance ρ_{ij}. $\Sigma_{\mathbf{\Theta}}$ is also a variable of the energy. In practice it can be computed from the Hessian of the energy. Compute the Hessian of the energy with respect to $\mathbf{\Theta}$ for a certain value of $\Sigma_{\mathbf{\Theta}}$, and obtain a new estimate of $\Sigma_{\mathbf{\Theta}}$. The computation of the parameter estimate is based on a three step iterative process: (i) minimize $E(\mathbf{\Theta},\Sigma_{\mathbf{\Theta}})$ with respect to $\mathbf{\Theta}$ holding the weights M_{ij} and the covariance matrix $\Sigma_{\mathbf{\Theta}}$ fixed. (iii) minimize $E(\mathbf{\Theta},\Sigma_{\mathbf{\Theta}})$ with respect to $\Sigma_{\mathbf{\Theta}}$ holding the weights M_{ij} fixed. (iii) Recompute the weight M_{ij} keeping $\Sigma_{\mathbf{\Theta}}$ fixed. Consequently, the full registration algorithm is also introducing a minimization step with respect to $\mathbf{\Theta}$, the weights and $\Sigma_{\mathbf{\Theta}}$ being fixed.

Inspired by these works we aim to recover uncertainties on the parameter vector $\mathbf{\Theta}$ defining the registering transformation as a covariance matrix corresponding to the degrees of freedom of the transformation. We have developed two different approaches to compute these uncertainties

leading to different results. We will show that these two approaches bears important similarity in the case of our study.

2.5.2 Continuous Formulation: Towards Hessian Matrix

Our approach relies on the linearity of the transformations. Therefore it could be used to handle uncertainty on any class of solutions to the approximation scheme presented in section 2.3.2. In particular, this is the case of Thin Plate Spline and Free Form Deformation which may be written with the form:

$$\mathcal{L}(\mathbf{\Theta}, \mathbf{x}) = \mathbf{x} + \mathcal{X}(\mathbf{x})\mathbf{\Theta}. \tag{2.44}$$

We are considering the quality of the local registration on shapes, that is the zero level set of the distance transform. Therefore, E_α is formulated in the limit case where α the size of the limited band around the model shape tends to 0. The data term of the energy function (Eq. 2.34) can now be expressed as:

$$E_0(\mathbf{\Theta}) = \int_{\partial\mathcal{S}} {\phi_\mathcal{T}}^2(\mathcal{L}(\mathbf{\Theta}; \mathbf{x}))d\mathbf{x} = \int_{\partial\mathcal{S}} {\phi_\mathcal{T}}^2(\mathbf{x}')d\mathbf{x}, \tag{2.45}$$

where we denote $\mathbf{x}' = \mathcal{L}(\mathbf{\Theta}; \mathbf{x})$. Consider $\mathbf{q}$ to be the closest point from $\mathbf{x}'$ located on $\mathcal{S}$, and given $\|\nabla\phi_\mathcal{T}(\mathbf{x}')\| = 1$ for a Euclidean distance, one can express the values of $\phi_\mathcal{T}(\mathbf{x}')$:

$$\phi_\mathcal{T}(\mathbf{x}') = \|\mathbf{x}' - \mathbf{q}\| = (\mathbf{x}' - \mathbf{q})\nabla\phi_\mathcal{T}(\mathbf{x}') \tag{2.46}$$

Considering little variation of the vector $\delta\mathbf{\Theta}$, one can express a first order approximation of the distance to the target:

$$\begin{aligned}
\phi_\mathcal{T}(\mathcal{L}(\mathbf{\Theta} + \delta\mathbf{\Theta}, \mathbf{x})) =& \phi_\mathcal{T}(\mathbf{x} + \mathcal{X}(\mathbf{x})\mathbf{\Theta} + \mathcal{X}(\mathbf{x})\delta\mathbf{\Theta}) \\
=& \phi_\mathcal{T}(\mathbf{x}') + \nabla\phi_\mathcal{T}(\mathbf{x}')^T.\mathcal{X}(\mathbf{x}).\delta\mathbf{\Theta} + \delta\mathbf{\Theta}^T\mathcal{X}(\mathbf{x})^T\mathcal{H}_\mathbf{\Theta}\mathcal{X}(\mathbf{x})\delta\mathbf{\Theta} \\
=& \nabla\phi_\mathcal{T}(\mathbf{x}')^T(\mathbf{x}' - \mathbf{q} + \mathcal{X}(\mathbf{x}).\delta\mathbf{\Theta}) + O(\|\delta\mathbf{\Theta}\|^2)
\end{aligned} \tag{2.47}$$

This local expression of $\phi_\mathcal{T}$ with a dot product reflects the condition that a point to curve distance was adopted. When located around a local minimum of E_0 first order terms are null, and we can

write the classical second order approximation of the quadratic energy in the form:

$$\begin{aligned} E_0(\boldsymbol{\Theta}+\delta\boldsymbol{\Theta}) &= \int_{\partial\mathcal{S}} \left[\nabla\phi_{\mathcal{T}}(\mathbf{x}')^T.(\mathbf{x}'-\mathbf{q}+\mathcal{X}(\mathbf{x}).\delta\boldsymbol{\Theta}) + O(\|\delta\boldsymbol{\Theta}\|^2)\right]^2 d\mathbf{x} \\ &= E_0(\boldsymbol{\Theta}) + \int_{\partial\mathcal{S}} \big[\delta\boldsymbol{\Theta}^T(\mathcal{X}(\mathbf{x})^T\nabla\phi_{\mathcal{T}}(\mathbf{x}')\nabla\phi_{\mathcal{T}}(\mathbf{x}')^T\mathcal{X}(\mathbf{x}))\delta\boldsymbol{\Theta}^T + \\ &\quad 2(\mathbf{x}'-\mathbf{q})^T(\nabla\phi_{\mathcal{T}}(\mathbf{x}')\nabla\phi_{\mathcal{T}}(\mathbf{x}')^T\mathcal{X}(\mathbf{x})^T)\delta\boldsymbol{\Theta}^T + 2(\mathbf{x}'-\mathbf{q})^T\nabla\phi_{\mathcal{T}}(\mathbf{x}')O(\|\delta\boldsymbol{\Theta}\|^2)\big]\, d\mathbf{x} \end{aligned} \tag{2.48}$$

This expression is satisfied near the optimum of the registration and therefore does not have first order terms in its expansion. Also a common assumption based on the Gauss-Newton method states that near the optimum, the term $(\mathbf{x}'-\mathbf{q})^T\nabla\phi_{\mathcal{T}}(\mathbf{x}')$ is on average close to zero. Consequently the third term in the above expression is negligible when compared to the first so that an efficient second order approximation is expressed as:

$$E_0(\boldsymbol{\Theta}_{min}+\delta\boldsymbol{\Theta}) = E_0(\boldsymbol{\Theta}_{min}) + \int_{\partial\mathcal{S}} \left[\delta\boldsymbol{\Theta}^T(\mathcal{X}(\mathbf{x})^T\nabla\phi_{\mathcal{T}}(\mathbf{x}')\nabla\phi_{\mathcal{T}}(\mathbf{x}')^T\mathcal{X}(\mathbf{x}))\delta\boldsymbol{\Theta}^T\right] \tag{2.49}$$

Localizing the global minimum of an objective function E is equivalent to finding the major mode of a random variable with density proportional to $\exp(-E/\beta)$. The coefficient β corresponds to the allowable variation in the energy value around the minimum. In the present case of a quadratic energy (and therefore Gaussian random variable), the covariance and the Hessian of the energy are directly related by $\Sigma_{\boldsymbol{\Theta}}^{-1} = H_{\boldsymbol{\Theta}}/\beta$. Having a quadratic form for the energy E_0 leads to the following expression for the covariance:

$$\Sigma_{\boldsymbol{\Theta},1}^{-1} = \frac{1}{\beta}\int_{\partial\mathcal{S}} \mathcal{X}(\mathbf{x})^T.\nabla\phi_{\mathcal{T}}(\mathbf{x}').\nabla\phi_{\mathcal{T}}(\mathbf{x}')^T.\mathcal{X}(\mathbf{x})d\mathbf{x} \tag{2.50}$$

In the most general case one can claim that the Hessian matrix $H_{\boldsymbol{\Theta}}$ is not invertible because the registration problem is under-constrained. This is in particular the case with FFD, as control points located too far from the zero level set have no impact on E_0. Then, additional constraints have to be introduced towards the estimation of the covariance matrix on $\boldsymbol{\Theta}$ through the use of a regularization term ($E_{smooth} = \boldsymbol{\Theta}^T\mathbf{C}\boldsymbol{\Theta}$) as presented in section 2.3.2. Such regularisation term presents the advantage of being invariant to affine transformations, it has an influence mostly on the data term of uncertainty estimation and consequently much weaker than a classical elastic regularizing term: Regularization constraints are often considered in computational vision to address ill-posed problems [194]. Towards an explicit formulation of the covariance on Θ, we consider such a term :

$$E(\mathbf{\Theta}) = \int_{\partial\mathcal{S}} \left[(\mathbf{x} + \mathcal{X}(\mathbf{x})\mathbf{\Theta} - \mathbf{q}) \cdot \nabla\phi_{\mathcal{T}}(\mathbf{x}')\right]^2 d\mathbf{x} + \lambda\, \mathbf{\Theta}^T\mathbf{C}\mathbf{\Theta} \tag{2.51}$$

This leads to the covariance matrix for the parameter estimate :

$$\Sigma_{\mathbf{\Theta},1} = \beta\left(\int_{\partial\mathcal{S}} \mathcal{X}(\mathbf{x})^T.\nabla\phi_{\mathcal{T}}(\mathbf{x}')\nabla\phi_{\mathcal{T}}(\mathbf{x}')^T\mathcal{X}(\mathbf{x})d\mathbf{x} + \lambda\mathbf{C}\right)^{-1} \tag{2.52}$$

This expression is well defined for arbitrarily small values of λ. When the value of λ tends towards 0, some eigenvalues of $\Sigma_{\mathbf{\Theta}}$ will diverge leading infinite uncertainty on the parameter vector $\mathbf{\Theta}$ along these directions. This will be further developed with the following approach that is based on the propagation of a Gaussian error on the data to the vector of parameters, leading to a slightly different result.

2.5.3 Discrete Formulation: Uncertainty propagation

There exists a different interpretation obtained when considering the variations of the error measurement near the optimum. In order to interpret and understand such a measure, let us consider the discrete case. Then, the objective function to be optimized is given by E_0:

$$E_0(\Theta) = \sum_{i=1}^{K} \phi_{\mathcal{T}}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x}_i))^2, \tag{2.53}$$

where $\{\mathbf{x}_i\}$ is a set of points uniformly sampled on the source shape boundary. One can also write the same type of second order approximation leading to the following quadratic form:

$$E_0(\mathbf{\Theta}) = \sum \left[(\mathcal{L}(\mathbf{\Theta}, \mathbf{x}_i) - \mathbf{q}_i)\nabla\phi_{\mathcal{T}}(\mathbf{x}_i')\right]^2,$$

where $\mathbf{x}_i' = \mathcal{L}(\mathbf{\Theta}, \mathbf{x}_i)$. Considering this expression in the neighborhood of the optimal parameter vector $\mathbf{\Theta}_{min}$ and the linear form of the transformation:

$$E_0(\mathbf{\Theta}_{min} + \delta\mathbf{\Theta}) = \sum \left[(\mathcal{X}(\mathbf{x}_i).\delta\mathbf{\Theta} + \underbrace{\mathbf{x}_i + \mathcal{X}(\mathbf{x}_i).\mathbf{\Theta}_{min}}_{\mathbf{x}_i'} - \mathbf{q}_i)\underbrace{\nabla\phi_{\mathcal{T}}(\mathbf{x}_i')}_{\eta_i}\right]^2$$

This is expressed in the compact matrix form:

$$E_0(\mathbf{\Theta}_{min} + \delta\mathbf{\Theta}) = (\hat{\mathcal{X}}\delta\mathbf{\Theta} - \mathbf{y})^T(\hat{\mathcal{X}}\delta\mathbf{\Theta} - \mathbf{y}), \tag{2.54}$$

with

$$\hat{\mathcal{X}} = \begin{pmatrix} \eta_1^T \mathcal{X}(\mathbf{x}_1) \\ \vdots \\ \eta_K^T \mathcal{X}(\mathbf{x}_K) \end{pmatrix} \text{ and } \mathbf{y} = \begin{pmatrix} \eta_1^T (\mathbf{x}'_1 - \mathbf{q}_1) \\ \vdots \\ \eta_K^T (\mathbf{x}'_K - \mathbf{q}_K) \end{pmatrix}, \tag{2.55}$$

where $\hat{\mathcal{X}}$ is a $K \times N$ matrix.We assume that $\mathbf{y}$ is the only random variable as it refers to the distance from the transformed source boundary $\delta\mathcal{S}$ to the target boundary $\delta\mathcal{T}$. Such assumption is equivalent with stating that errors in the point positions are only quantified along the normal direction. This accounts for the fact that the point set is treated as samples extracted from a continuous manifold. One can take the derivative of the objective function in order to recover a linear relation between Θ and $\mathbf{y}$:

$$\hat{\mathcal{X}}^T \hat{\mathcal{X}} \delta\mathbf{\Theta} = \hat{\mathcal{X}}^T \mathbf{y} \tag{2.56}$$

The set of values contained in $\mathbf{y}$ refers to the point to curve distance between the transformed source and the target. Therefore one can assume that the components of $\mathbf{y}$, measuring the local registration error, are a set of independent and identically distributed random variables. A basic assumption will use for $\mathbf{y}$ a multivariate Gaussian random variable. The covariance matrix of $\mathbf{y}$ has the form $\sigma^2\mathbf{I}$ of magnitude σ^2 with $\mathbf{I}$ being the identity matrix. Due to the linear form, the Gaussian assumption for $\mathbf{y}$ leads to a Gaussian random variable for $\mathbf{\Theta}$. Once again, in the most general case one can claim that the matrix $\hat{\mathcal{X}}^T\hat{\mathcal{X}}$ is not invertible due to the fact that the registration problem is under-constrained. In a similar fashion, this problem is addressed adding the regularization term $E_{smooth} = \mathbf{\Theta}^T\mathbf{C}\mathbf{\Theta}$:

$$(\hat{\mathcal{X}}^T \hat{\mathcal{X}} + \lambda C)\delta\mathbf{\Theta} = \hat{\mathcal{X}}^T \mathbf{y} \tag{2.57}$$

Therefore the Gaussian random variable $\mathbf{\Theta}$ is fully determined for any arbitrarily small value of λ. $\mathbf{\Theta}$ has covariance:

$$\Sigma_{\mathbf{\Theta},2} = \sigma^2(\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda\mathbf{C})^{-1}\hat{\mathcal{X}}^T\hat{\mathcal{X}}(\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda\mathbf{C})^{-1} \tag{2.58}$$

This expression is different from $\Sigma_{\mathbf{\Theta},1}$ which in the discrete case would simply be written:

$$\Sigma_{\mathbf{\Theta},1} = \beta(\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda\mathbf{C})^{-1} \tag{2.59}$$

The next section will explore the relations between the two expressions.

2.5.4 Relations between the Hessian approach and Uncertainty propagation

Results obtained within the two approaches are fundamentally different since in the first case, we are measuring the Hessian of the energy function while in the second case, modifying the data and

capturing the variation of the minimum of the energy. In particular, $\Sigma_{\mathbf{\Theta},2}$ is a singular matrix (as opposed to $\Sigma_{\mathbf{\Theta},1}$). This means that not all of the displacements in the parameter space $\mathbf{\Theta}$ can be explained with variations of the data. In the formula, this is due to the fact that the matrix $\mathcal{X}$ is not full rank.

We will give some hints on how these two approaches are related by considering a vector of the null space of $\Sigma_{\mathbf{\Theta},2}$ which will be denoted by $Ker(\Sigma_{\mathbf{\Theta},2})$. Given the form of $\Sigma_{\mathbf{\Theta},2}$ and the fact that $(\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \gamma\mathbf{C})$ is invertible, any vector of $Ker(\Sigma_{\mathbf{\Theta},2})$ is related to the null space of $\hat{\mathcal{X}}^T\hat{\mathcal{X}}$. Consider a vector $\mathbf{v}_1$ verifying $\hat{\mathcal{X}}^T\hat{\mathcal{X}}.\mathbf{v}_1 = 0$. Consider the singular value decomposition of $\mathcal{X}$ and assume it is of rank R:

$$\hat{\mathcal{X}} = U_{K,R}.D_R.V_{N,R}^T, \tag{2.60}$$

with $U_{K,R}$, $V_{N,R}$ being matrices with orthonormal column vectors and D_R a diagonal matrix with all non-zero values. When applied to $\mathbf{v}_1$:

$$\begin{aligned}
&\hat{\mathcal{X}}^T\hat{\mathcal{X}}.\mathbf{v}_1 = V_{N,R}.D_R^2.V_{N,R}^T.\mathbf{v}_1 = 0 \\
&\Rightarrow V_{N,R}^T.\mathbf{v}_1 = 0 \\
&\Rightarrow \hat{\mathcal{X}}.\mathbf{v}_1 = 0 \\
&\Rightarrow \forall i, \eta_i^T\mathcal{X}(\mathbf{x}_i).\mathbf{v}_1 = 0.
\end{aligned}$$

This corresponds to directions along the vector of parameter that either do not influence contour points ($\mathcal{X}(\mathbf{x}_i).\mathbf{v}_1 = 0$) or influence contour points in the normal direction ($\eta_i^T\mathcal{X}(\mathbf{x}_i).\mathbf{v}_1 = 0$). From this point, the influence on the associated directions of the covariance matrix is inferred :

$$\begin{aligned}
&\Sigma_{\mathbf{\Theta},2}.\mathbf{v}_0 = 0 \\
&\Rightarrow (\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda.C)^{-1}.\hat{\mathcal{X}}^T\hat{\mathcal{X}}.(\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda.C)^{-1}.\mathbf{v}_0 = 0 \\
&\Rightarrow \hat{\mathcal{X}}^T\hat{\mathcal{X}}.\underbrace{(\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda.C)^{-1}.\mathbf{v}_0}_{\mathbf{v}_1} = 0 \\
&\Rightarrow \mathbf{v}_0 = (\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda.C).\mathbf{v}_1 \\
&\Rightarrow \mathbf{v}_0 = \lambda C\mathbf{v}_1.
\end{aligned}$$

Based on this equation, any vector of $Ker(\Sigma_{\mathbf{\Theta},2})$ can be written as the product of the null space of $\hat{\mathcal{X}}^T\hat{\mathcal{X}}$ with the smoothing matrix C. These two vector space have the same dimension because the matrix C is built in such a way that the matrix $\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda.C$ is non-singular.

In the particular case where C is proportional to the identity matrix ($C = \gamma I$), corresponding to a purely elastic smoothing term, we have $\mathbf{v}_1 \propto \mathbf{v}_0$ so that $\mathbf{v}_0$ is also an eigenvector of $\Sigma_{\mathbf{\Theta},1}$ with eigenvalue $1/(\lambda\gamma)$. This value corresponds to the largest direction of uncertainty.

We may consider the general form for the matrix C in the limit case when the smoothing term is arbitrarily small: let $\mathbf{v}_0$ be a vector of $Ker(\Sigma_{\mathbf{\Theta},2})$, there exists $\mathbf{v}_1$ verifying $\mathbf{v}_0 = C\mathbf{v}_1$, with $\mathbf{v}_1$ an element of the null space of $\hat{\mathcal{X}}^T\hat{\mathcal{X}}$, we also have the relation:

$$\mathbf{v}_0 = \frac{1}{\lambda}(\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda C)\mathbf{v}_1 \tag{2.61}$$

Applying it to the Hessian based expression of uncertainty $\Sigma_{\mathbf{\Theta},1}$: $\mathbf{v}_0\Sigma_{\mathbf{\Theta},1}\mathbf{v}_0 = \frac{1}{\lambda}\mathbf{v}_0\mathbf{v}_1 = \frac{1}{\lambda}\mathbf{v}_0 C\mathbf{v}_0$. In such cases, the vectors of $Ker(\Sigma_{\mathbf{\Theta},2})$ correspond to directions so that $\mathbf{v}_0\Sigma_{\mathbf{\Theta},1}\, bv_0$ goes to infinity (the norm of $\Sigma_{\mathbf{\Theta},1}.\mathbf{v}_0$ is diverging) when the smoothing term is negligible ($\lambda \rightarrow 0$). As our framework considers a coarse to fine registration process where the influence of the regularization term tends to zero; the above condition is satisfied.

This means that the vectors of $Ker(\Sigma_{\mathbf{\Theta},2})$ are excluded (null probability) if using uncertainty based on the variations of data, while these directions are irrelevant (nearly as likely as the optimal parameters) if using uncertainty based on the Hessian of the energy.

Once the behavior of $\Sigma_{\mathbf{\Theta},1}$ and $\Sigma_{\mathbf{\Theta},2}$ has been identified on $Ker(\Sigma_{\mathbf{\Theta},2})$, one may consider its orthogonal space ($Ker(\Sigma_{\mathbf{\Theta},2})^{\perp}$). Consider a vector $\mathbf{y}$ of the space $Ker(\Sigma_{\mathbf{\Theta},2})^{\perp}$ and form the product $\Sigma^{-1}_{\mathbf{\Theta},1}\Sigma_{\mathbf{\Theta},2}$:

$$\Sigma^{-1}_{\mathbf{\Theta},1}\Sigma_{\mathbf{\Theta},2}\mathbf{y} = \mathbf{z} \tag{2.62}$$

$$\hat{\mathcal{X}}^T\hat{\mathcal{X}}\hat{\mathbf{y}} = (\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda.C)\hat{\mathbf{z}} \qquad \text{with} \qquad \begin{cases} \mathbf{y} = (\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda.C)\hat{\mathbf{y}} \\ \mathbf{z} = (\hat{\mathcal{X}}^T\hat{\mathcal{X}} + \lambda.C)\hat{\mathbf{z}} \end{cases} \tag{2.63}$$

$$\hat{\mathcal{X}}^T\hat{\mathcal{X}}(\hat{\mathbf{y}} - \hat{\mathbf{z}}) = \lambda.C.\hat{\mathbf{z}} \tag{2.64}$$

Let consider the limit case when the regularization becomes negligible, we obtain the identity between $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ on $Ker(\Sigma_{\mathbf{\Theta},2})^{\perp}$.

Therefore the present approach based on the variations of the data (leading to $\Sigma_{\mathbf{\Theta},2}$) can be seen as a mean to remove the influence of the variations of the parameter vector which do not affect the data. The choice of the uncertainty model ($\Sigma_{\mathbf{\Theta},1}$ or $\Sigma_{\mathbf{\Theta},2}$) should depend on the application. It mostly depends whether or not the null space associated with the registration has to be considered. In the present case, uncertainty will be used to express the posterior density of the parameter vector, given the registration. Consequently, $\Sigma_{\mathbf{\Theta},1}$ will be preferred so that density is defined on the entire space.

In practice the calculation of the covariance is not necessary as it will be used to express the registration likelihood through Gaussian kernels. So the uncertainty on the parameters of the

registration will be simply expressed by the inverse of the covariance matrix:

$$\Sigma_{\mathbf{\Theta}}^{-1} = \sigma^{-2}(\mathcal{X}^T\mathcal{X} + \gamma\mathbf{C}) \tag{2.65}$$

Furthermore the sums and integrals presented in equations 2.58 and 2.52 do not consider a discretization of contour or surface but make use of a set of points uniformly sampled in space and located in a thin narrow band surrounding the source shape/contour.

2.5.5 Scaling Uncertainty

Within both expressions of the uncertainty matrix a scaling coefficient appears (β in eq. 2.52, σ^2 in eq. 2.58). We are retrieving this value in a heuristic manner based on the physical interpretation of the Energy. For small displacements of the parameter vector close to the optimum of the energy we have the relation:

$$dE = \int_{\partial\mathcal{S}} {\phi_{\mathcal{T}}}^2(\mathcal{L}(\mathbf{\Theta}_{min} + d\mathbf{\Theta}; \mathbf{x}))d\mathbf{x} - E_0(\mathbf{\Theta}_{min}) = \frac{1}{2}d\mathbf{\Theta}.H.d\mathbf{\Theta}. \tag{2.66}$$

Within our application, we have scaled the Energy so that the hyperellipsoid boundary defined by the implicit equation $d\mathbf{\Theta}.H.d\mathbf{\Theta} = 1$ has a physical meaning when applied to the shape. The purpose of uncertainty will be to represent a single shape as a distribution of possible deformations, mostly using a Gaussian random variable with a covariance matrix $\Sigma_{\mathbf{\Theta}}$ being related to $H_{\mathbf{\Theta}}^{-1}$. In order to estimate the scale of uncertainty, we will consider a special case where every control points are fixed except one. Displacement of a control point can be easily related to a local displacement of the shape boundary, and therefore directly related to the variation of the energy.

In the case of TPS, we have considered that all control points have equal influence on the surface to be transformed. Consequently the we have considered the total area of the shape boundary divided by the number of control points as an influential area which we denote $\mathcal{A}_{\mathbf{P}}$. In the case of $3D$ FFD, we have considered the squared distance between control points. This quantity is the influential area of any control points $\mathcal{A}_{\mathbf{P}}$.

If an additional normal motion of the surface is allowed within a distance d_0 equal to 1% of a charasteristic size of the shape, we obtain a quantity homogeneous to the energy corresponding to the 'allowable' range of displacements of the shape boundary that is used for scaling:

$$E_{scale} = \mathcal{A}_{\mathbf{P}}.d_0^2 \tag{2.67}$$

Therefore we have used in our applications the uncertainty matrix the quantity $(E_{scale}.H_{\mathbf{\Theta}}^{-1})$. This uncertainty matrix is displayed with the form of ellipses and ellipsoids, each ellipsoid corresponds to the case where every other control point remain fixed at their optimal position. If using Gaussian random variables to represent the distribution of displacements, these ellipses in $2D$ concentrate 40% of the total probability while the ellipsoids in $3D$ concentrate 20% of the probability.

2.5.6 Experimental Results

Some examples of such estimates are shown in Figure 2.12 and 2.13 where in the 2D case, 2×2 diagonal blocs of the $2N \times 2N$ uncertainty matrices $\Sigma_{\mathbf{\Theta}}^{-1}$ are extracted, and represented as a set of ellipses located at every control points corresponding to the loci of $\mathbf{x}$ verifying $\mathbf{x}^T.\Sigma_{\mathbf{\Theta},22}^{-1}.\mathbf{x} = 1$. One should notice that such representation is considering the control points as being independent and do not represent all the covariant factors existing between control points. These coefficient are not negligible and play an important role on neighboring control points. Figure 2.12 shows uncertainty ellipses applied to FFD. Ellipses are elongated in the direction of the contour/surface and larger on control points located further away from the shape boundary, which is due to the fact that these control points have smaller influence on the shape deformations. Figure (2.13) and

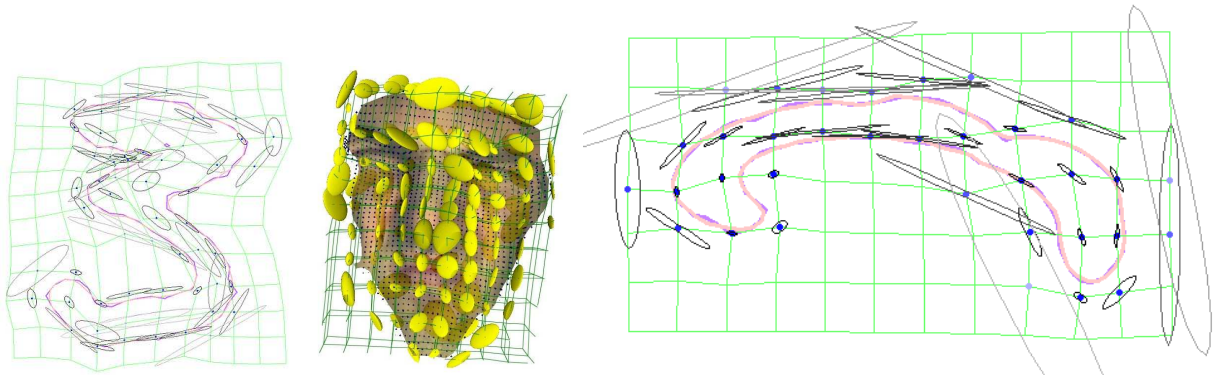
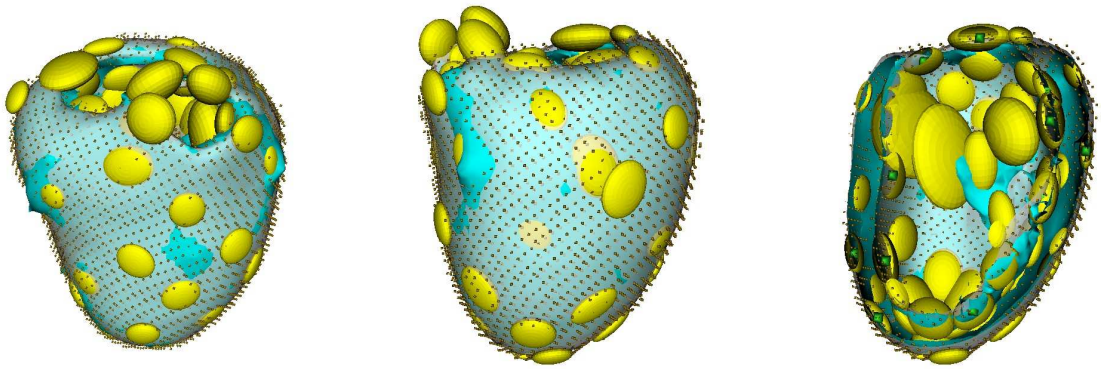


Fig. 2.12: Projection of the covariance matrix Σ_{Θ}^{-1} on the grid control points. Ellipses are elongated in the direction of the contour, and larger on distance control points. Ellipses are not represented at control points where uncertainty is too large.

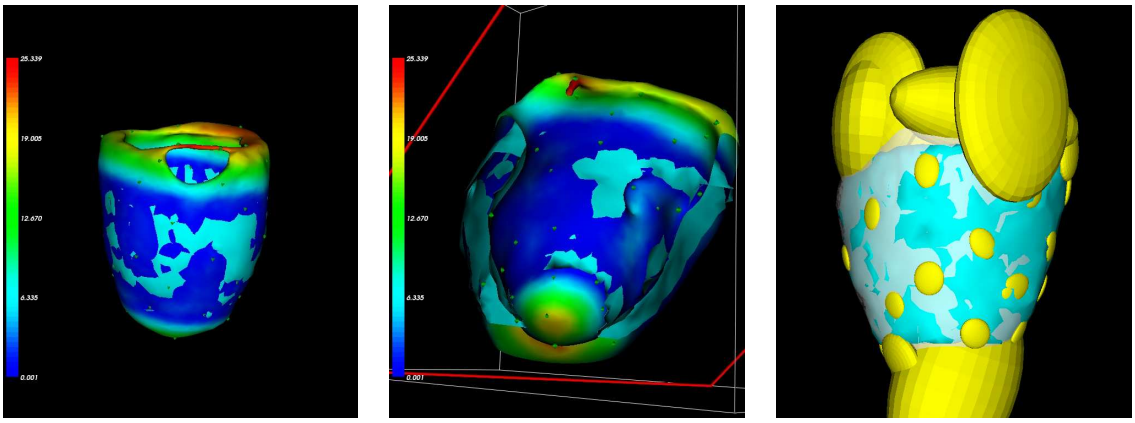
(2.14) shows uncertainty ellipsoids applied to TPS transformation. control points being located on the shape boundary, they have equal influence on the shape. Ellipsoids are elongated in the direction of the of the surface and shows larger uncertainty at locations where the match is not exact. One notice such larger uncertainty inside the ventricle near the papillary muscles. This is due to the fact that the automatically segmented left ventricle present very important variations in

this area (Fig. 2.13). In the case of incomplete MRI data (Fig. 2.14), uncertainty is extremly large on the missing parts of the target shape.



(a) (b) (c)

Fig. 2.13: Projection of the covariance matrix Σ_Θ on the TPS control points. (a, b) Left ventricle epicardium, (c) Left ventricle endocarium (same pose as (b)).



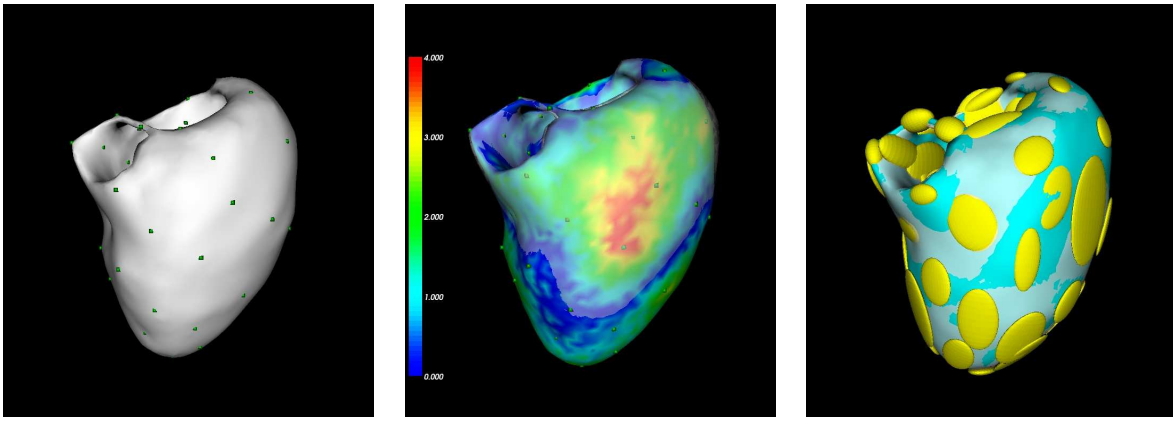
(a) (b) (c)

Fig. 2.14: Registration and uncertainty computation on a manually segmented ventricle from MRI data. Colormap shows the registration error (in mm.), (a) full view, (b) clipped view. (c) Uncertainty ellipses.

2.5.7 Testing Sequence

In order to test the relevance of the registration framework and the computation of uncertainty, we have carried a special experiment based on the training set. We dispose of a training set of left ventricular shapes extracted from Computer Tomography imaging. A reference shape of the

left ventricle has been registered to every sample of the training set using a Thin Plate Spline Transformation with 90 control points (let $\mathbf{\Theta}_i$ be the parameter vector of such transformation). For every registered shapes, uncertainty has been computed with the above method (let $\Sigma_{\mathbf{\Theta}_i}$ be the associated covariance matrix). Then new random transformations are created, sampled from a Gaussian distribution, centered on the transformation with parameter vector $\mathbf{\Theta}_i$ and covariance proportional to the inverse of uncertainty ($\Sigma^{-1}_{\mathbf{\Theta}_i}$). Consequently the perturbations appears mostly in the normal direction to the surface. 75 random transformations and their associated shapes are generated using this method. The reference shape model is registered to the set of perturbed samples (vector $\mathbf{\Theta}_{i,reg}$) and uncertainty is compute for every registration result.



(a) (b) (c)

Fig. 2.15: (a) An example of the registration output: from the reference model to a shape of the training set ($\mathbf{\Theta}_i$). (b) This registration output with an additional random perturbation ($\mathbf{\Theta}_{i,perturb}$). (c) Registration of the reference model on the perturbed sample (transformation $\mathbf{\Theta}_{i,reg}$)+ uncertainty representation ($\Sigma_{i,reg}$).

Figure 2.15 shows image example of the registration output on a training set example, addition perturbation and the registration result on the example with perturbation. A set of measure can be performed to assess the importance of uncertainty in the vector space of the Thin Plate Spline. The table 2.16 represents a set of measures based on the control points. The average distance between the perturbed sample and the initial one ($\|\mathbf{\Theta}_i - \bar{\mathbf{\Theta}}_{i,perturb}\|^2$) is lower than the average distance between the registered sample and the perturbed one ($\|\mathbf{\Theta}_{i,reg} - \bar{\mathbf{\Theta}}_{i,perturb}\|^2$). This situation is not acceptable as the registered sample is designed to be closer to the perturbed one. The use of uncertainty information, allow to compute efficiently distance between shapes directly in the vector space of TPS transformations.

	mean	std deviation
$\|\mathbf{\Theta}_{i,reg} - \mathbf{\Theta}_{i,perturb}\|^2$	23.70	4.69
$\|\mathbf{\Theta}_{i} - \mathbf{\Theta}_{i,perturb}\|^2$	20.13	1.53
$\log(N(\mathbf{\Theta}_{i,reg} - \mathbf{\Theta}_{i,perturb}, \Sigma_i))$	-694.43	6.63
$\log(N(\mathbf{\Theta}_{i,reg} - \mathbf{\Theta}_{i,perturb}, \Sigma_i))$	-721.36	14.53

Fig. 2.16: This table shows that computing distances in the vector space of the transformation may only be relevant if uncertainty information is also used. With the use of uncertainty, the distance between the perturbed sample and its registration is always smaller than the distance between the perturbed sample and the same sample without perturbation.

2.6 Conclusions

We have presented in this chapter the multiple representations of shapes and their relations to different problems of computer vision. Independently, we have presented spatial approximation/interpolation techniques, that lead to a linear representation of spacial deformations. We have focused on two particular cases Free Form Deformation and Thin Plate Splines and combined them with a shape representation based on distance transforms.

The main contributions of our work lie in two particular aspects. Although combining spatial transformation and distance representation has been previously explored, we have pushed the optimization process one step further to handle large deformations. To do this we had to address the problems of narrow bands with variable size and different sparseness upon discretization. Parallel to this, increasing the degrees of freedom of the underlying deformation allows the registration process to go from a rough shape alignment to warping while preserving an invertible transformation. Our second contribution lies in the estimation of uncertainties in the registration result. Due to the parametric transformation and noise of the objective shape, local registration cannot lead to an exact warping of shapes. Assessing uncertainty at the parameter level can be of greater importance than the local error as it represents the allowable range of variation of the parameter leading to similar registration results.

Future research directions may cover different perspectives. First of all, different shape representations sharing similarity invariant properties can be tested like integral kernels. Regarding the optimization problem, it would be of interest to adapt the sparseness of the sampling locally to address more efficiently the redundancy of the distance transform, therefore sampling denser points in areas with high curvature of the iso-levels representing the details of the shape.

Furthermore, the uncertainty information in section 2.5, are presently computed on the shape

boundary of the final registration result. This can be extended to different iso-contours and used to select a different model of transformations or refine the narrow band.

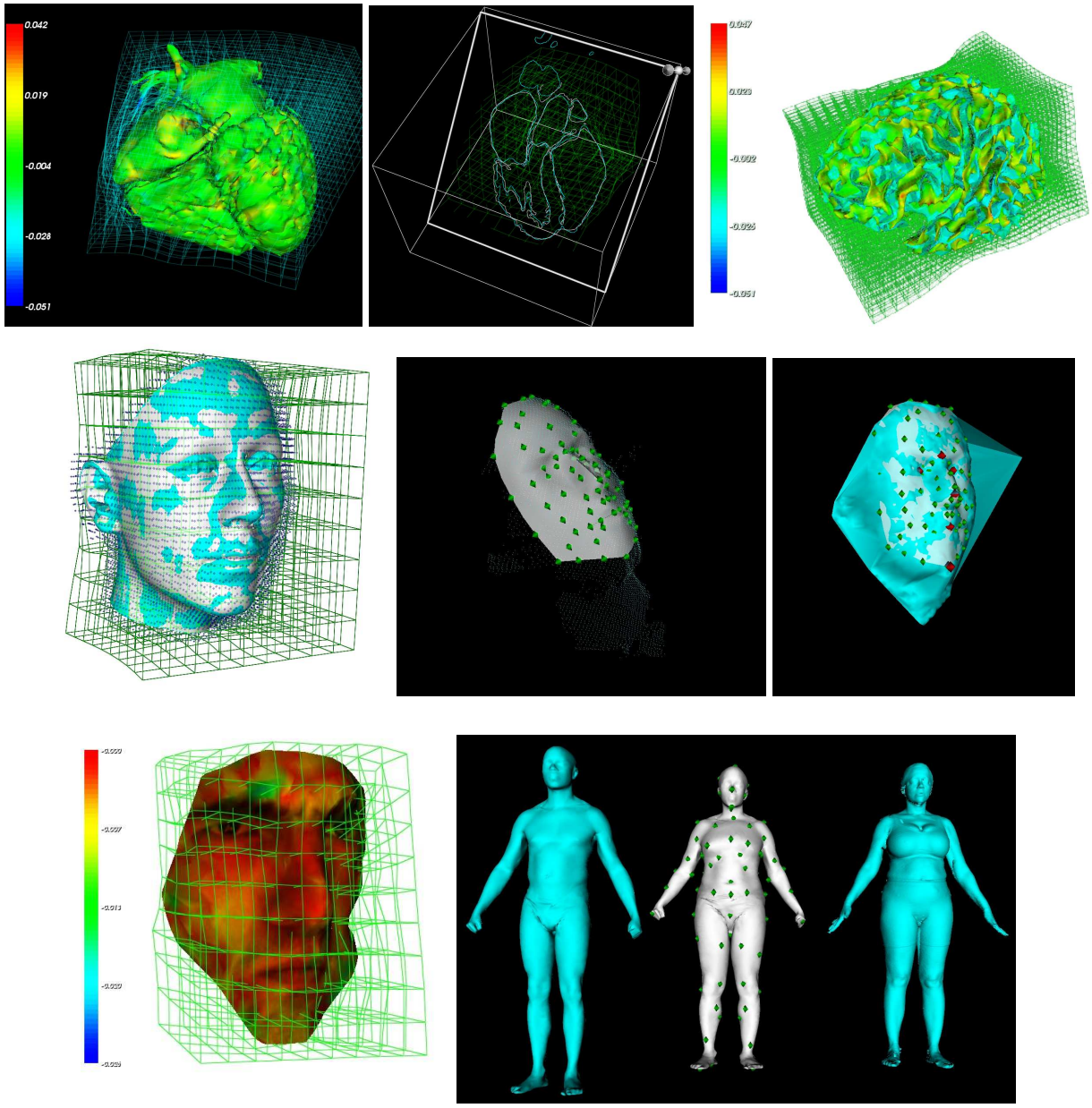


Fig. 2.17: Various samples of registration

Chapter 3

Modeling Shape Variations

Abstract – *A definition of shape modeling is related to the modeling of random variables defining the variations in a class of shapes. Shape being an abstract concept, shape modeling is also naturally related to the shape representation. This chapter will focus on statistical shape modeling that consists of a set of techniques that can be applied to describe the possible variations of a class of shapes of interest. Given a training set of shapes, considered as a set of random samples drawn from an unknown probability in a possibly infinite dimensional manifold, the main objective is to define the parameters of the density that expresses the observed variability with low complexity. A rich set of methods starting from the parametric ones, assuming an underlying form of the density and ending with the non-parametric ones based on kernels density estimation will be presented. In between these classes, one can find methods that aim first to transform the observations non-linearly to a linearly separable space and then use conventional parametric methods to model their variability. In this chapter we will focus on the use of uncertainty both in the case of parametric modeling with Independent Component Analysis and non-parametric modeling using variable bandwidth kernel density estimation.*

3.1 Introduction

Let us consider a random variable as well as some measurements/observations/samples coming from this variable. In the most general case these measurements are incomplete and unable to characterize the behavior of the variable. The task of recovering a statistical model describing this variable is a well-studied problem in statistics, pattern recognition and machine learning. The task can be mainly decomposed in two components, find the most appropriate statistical model to express the variable and estimate the parameters of this model to match the sparse measurements/observations.

The case of modeling shape variations falls in the above general definition. Once the registration of all training examples to the same pose is completed, the task of recovering a compact statistical model is to be addressed. The main challenge is to address on one hand the important dimensionality of the samples (3D shapes) while at the same time the small number of observations (number of manual segmentation being available). Therefore important attention is to be paid both to the selection of the model as well as to the inference approach on defining its parameters. On top of that, one should be able to account for the uncertainties determined during the registration phase, which makes the problem more complicated. This is due to the fact that each measurement (shape, random variable), consists of individual measurements which are also random variables and have to be propagated properly to the final model.

Prior art in modeling patterns from examples as well as shape variations exploits (i) Parameter estimation and supervised learning [31, 196], (ii) Non-parametric densities approximation [48, 150], (iii) linear discriminant functions [63, 102]. Parameter estimation through supervised learning makes an explicit assumption on the nature of the distribution which approximates the samples [176]. Then, the maximum likelihood principle is used to determine the parameters of the model. These models consider known parametric densities, like Gaussian [99] as well as mixtures of Gaussians [44]. The unknowns to be determined are the number of mixture components, their individual parameters (mean, variance, etc.) as well as the prior of each component. The use of Minimum Description Length [52], or the Schwartz criterion from Bayesian information theory are considered to determine the number of components while the Maximum Likelihood is used to determine the parameters of each distribution. In the most simple variant of this model - often considered to model shape variations -, the use of single-mode Gaussian density has been considered with their mean and variance being determined through singular value decomposition of the data matrix. These methods are a reasonable compromise between complexity and computational efficiency under the condition that the selected model can well express the samples (we assume that

the density from which the samples are taken is known). One has to determine a small number of variables from a reasonable observation set (well suited problem for unimodal distributions). On the other hand their estimation might be quite sensitive to the initial conditions for the individual parameters of the mixture components in particular for multi-variate densities with a large number of components. Finding the tradeoff between the approximation error and the cost of introducing new components (which eventually will improve the approximation quality) is not straightforward.

Opposite to parametric statistical modeling, the non-parametric density approximation approach does not make any assumptions on the underlying density. One should note that these methods are either used for modeling statistical behavior of samples or classification. In the former most common case, these methods aim to reproduce the entire density from a small number of observed/selected patterns. On the other hand, when aiming classification one can address the task through direct comparison between the new observations and the samples. Parzen windows are the most popular models to describe non-parametric densities. In such a context, one can consider approximation functions (often called window-functions) and then express the entire density through a linear combination of these window functions being centered at the observed samples. Gaussians are the most common choice for the window functions. The methods often rely on fixed bandwidth, where each sample is given the same importance. More advanced methods go beyond this simplistic assumption and introduce a general class of kernels (e.g. window functions) where the bandwidth (importance) of the sample function can vary across observations. In both cases, the selection of samples used to reproduce the density is crucial. K-nearest neighbor estimation is an alternative to kernel-based non-parametric density estimation aiming to properly address the samples selection. These methods are seeking k-clusters, or representative samples which for a given window function can approximate the surrounding cluster area given the whole observation set. These methods can offer efficient classification but one has to figure out appropriate means to determine the required number of clusters. Furthermore, their performance is questionable when seeking a complete statistical description of the samples, like in the case of medical image analysis and computer aided diagnosis. In the most general case one can claim that the use of non-parametric densities provides efficient tools to model shape variations. The principle difficulty of such approach lies in the sample selection (retain the most important ones) and the choice of data-driven kernels in terms of their corresponding bandwidth.

One of the main challenges in 3D statistical shape modeling is the large dimensionality of the samples. In the most general case even a simple surface representation with control points will require random variables in a space of 50 to 100 parameters. Statistical inference in these spaces with predefined models is not straightforward since in most of the cases the space is not linear while at the same time the use of non-parametric densities might be impractical. Dimensionality

reduction techniques have been investigated to address this problem. PCA which has been earlier presented as well as methods like non-negative matrix factorization have been investigated. Ideally one would like to reduce the dimensionality and introduce linearity to the new space. Such a task offers an alternative concept to statistical modeling, aiming more towards classification where one seeks discriminant functions to determine the bounds of the observed densities. Once the bounds have been determined, the answer to the classification problem is straightforward. In the context of our work we are interested in knowledge-based segmentation and therefore a need exists to convert statistical models to constraints. The use of bounds, discriminant functions and hyper-planes offers a natural way to constrain the segmentation problem. Discriminant functions can be either linear in the observations or linear in a space recovered through the projection of the observations according to some transfer function. Linear (Fisher) discriminant analysis [63] seeks a linear transformation matrix in such a way that the ratio of the between-class scatter and the within-class scatter is maximized. Independent Component Analysis defines a generative model where the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown [87]. The latent variables are assumed non-Gaussian and mutually independent and they are called the independent components of the observed data.

Machine learning methods go beyond simple linear classification techniques. The use of kernels to project the non-linear observations to a linear space and then the use of linear approximations in this new space is an active on-going research effort in the area of machine learning [160]. The most representative examples of this effort refers to Kernel-Principal Component Analysis [119]. These methods seek a set of operators/kernel functions which will project the original space onto a new linearly-separable space, where one can produce an accurate statistical approximation of the new samples (creation of clusters), or be able to separate the different hypotheses. The same concept was developed for the independent component analysis concept. Linear embedding is an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs. Support vector machines is a powerful clustering method which seeks an hyper-plane and a normal vector with the least possible norm which separates correctly the labeled data. Boosting methods like adaboost are examples of weak linear classifiers which upon proper integration can lead to exceptional classification performance. One should also point out recent work on compressed or compressive sensing [60] which consists of recovering from a set of subspaces the least possible number of examples capable of expressing the observations under a sparsity assumption.

The current state of the art in classification demonstrates that separation/clustering between samples can be efficiently done provided sufficient training data are available. The problem of segmentation though requires a probabilistic representation of the samples and in our case goes

beyond traditional density estimation due to the presence of uncertainty estimates. Furthermore, organs exhibit various degrees of variability which are to be considered. The main contribution in this chapter presents the design of a non-parametric density function approximation with variable bandwidth kernels. This density explores both the deformation and the uncertainties space where the kernel bandwidth is directly associated with the local uncertainties on the registration space. We also introduce a marginal variation of the independent component analysis, which introduces the use of uncertainty information on the training set for the computation of the independent components.

The remainder of this chapter is organized as follows. First, we review traditional dimensionality reductions techniques like principal and independent component analysis. The next section reviews parametric and non-parametric approximations for samples using fixed and variable bandwidth. Then, we introduce two novel statistical models which aim to propagate the uncertainties which are associated with the samples: (i) uncertainty-driven independent component analysis, and (ii) uncertainty-based variable bandwidth non-parametric density approximation. We conclude this chapter with potential future directions to further explore measurements of uncertainties associated with the samples towards more accurate statistical modeling of random variables.

3.2 State of the Art in Statistical Interpretation of Samples

Statistical analysis and interpretation of samples often involves input data drawn from one or more unknown random variables.
The main objectives of this process are:

- *Classification*: is used to label the data in a certain number of classes. An interpretation of the problem considers that data points associated with each class are sampled according to an unknown set of independent random variables. The training set being built with samples drawn from each class, the classification problem consists of estimating the conditional probability of a class given the data, and grants to any unclassified data point the label of the most likely class.

- *regression*: assuming that a relationship exists between two random variables, the regression problem consists of defining the conditional probability of one of the variable given the other.

- *density estimation*: assuming that the observed samples come from a single unknown random variable, the aim is to determine the support and the parameters of the corresponding density.

These tasks are often addressed in two stages involving a learning and an execution phase. During learning, samples, corresponding classes and dependencies are given and the aim is to determine the regressor. The definition of above tasks introduces certain relationships/dependencies between them: density estimation actually encompasses the two others as classification relies on the probability of the class given a certain sample, and regression the probability of one of the variable given the other. Therefore these application were addressed with the same kind of probabilistic models.

Let us consider a training set $\{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_\mathrm{N}\}$ of shapes representing the structure of interest. Our approach to shape registration requires that a common shape model be registered to every sample of the training set with respect to a certain parametric transformation. As a consequence, the training set is represented as a set of transformations:

$$\mathcal{H} = \{\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_N\},$$

where $\mathbf{\Theta}_k \in \mathbb{R}^d$. The statistical modeling task consists of recovering a probabilistic representation of this set. The task of building the common reference shape model will be developed in section (3.3), we focus in this section on most standard methods for density estimation. The first common assumption states that every $\mathbf{\Theta}_k$ is an independent sample drawn from an unknown random variable with density $p(\mathbf{\Theta})$.

3.2.1 Bayesian Learning

Let us consider without loss of generality the parametric form of the density is known, and it involves a set of parameters ω to be determined. The samples of $\mathcal{H}$ are therefore drawn independently according to the probability law $p(\mathbf{\Theta}|\omega)$, and therefore one can express the density of $\mathcal{H}$ conditional on ω:

$$p(\mathcal{H}|\omega) = \Pi_{k=1}^{n} p(\mathbf{\Theta}_k|\omega).$$

That is called the likelihood of the samples. The most common mean to determine the parameters ω is through the *maximum likelihood estimate* is the value $\hat{\omega}$ that maximizes $p(\mathcal{H}|\omega)$.

In some problems, on top of knowledge of the form of the density, we also retain some information on the distribution of the parameters ω. This can be considered as a prior density $p(\omega)$, and the optimal estimate of ω should be able to express the samples while corresponding to a peak on the prior density. In such a case, the probability density for ω given the samples is written

according to the Bayes rule:

$$\begin{aligned} p(\omega|\mathcal{H}) &= \frac{p(\mathcal{H}|\omega).p(\omega)}{\int p(\mathcal{H}|\omega).p(\omega)d\omega} \\ &\propto p(\mathcal{H}|\omega).p(\omega) \end{aligned}$$

Given that the denominator is independent to ω, the optimal solution for ω is the one that maximizes this density and is known as the *maximum a-posteriori*.
This is usually obtained with the minimization of its log likelihood $L(\omega|\mathcal{H}) = -[\log(p(\mathcal{H}|\omega)) + \log(p(\omega))]$. If there is no prior on ω, which can be modeled with a uniform probability density, maximum likelihood and maximum a-posteriori coincide. Normal distributions have certain nice theoretical properties and are often considered to model single and high-dimensional samples. Therefore a need exists to study this particular class of models which are often considered in computer vision and medical image analysis.

3.2.2 Gaussian Assumption

Now assume that the training sample is drawn from a Gaussian distribution. Within this assumption, the samples $\mathbf{\Theta}$ have to be located in a finite dimensional vector space, that is isomorphic to $\mathbb{R}^d$. Then, if the mean and variance are unknown variables, the maximum likelihood criterion is obtained solving its derivative with respect to the parameters:

$$\frac{d\left\{\sum_{k=1}^{k=N}\log(p(\mathbf{\Theta}_k|\mu,\Sigma))\right\}}{d(\mu,\Sigma)} = 0.$$

This leads to the following estimator [4]:

$$\hat{\mu} = \bar{\mathbf{\Theta}} = \frac{1}{N}\sum_{k=1}^{k=N}\mathbf{\Theta}_k \qquad (empirical\ mean)$$

$$\hat{\Sigma} = \frac{1}{N}\sum_{k=1}^{k=N}(\mathbf{\Theta}_k - \hat{\mu}).(\mathbf{\Theta}_k - \hat{\mu})^T \quad (empirical\ variance)$$

The most common objective for performing such an estimation is dedicated to the reduction of the problem dimensionality. One can determine the empirical mean and then subtract it from the samples in order to recover the empirical covariance matrix. The study of the eigenvectors of the covariance estimates leads to a representation in a different rotated basis of the samples

into which the components are uncorrelated. It also allows to capture modes of variation for the samples or determine the set of variations that are most often observed in the density. The eigen decomposition of the covariance is composed of a unit matrix U and positive diagonal matrix Λ:

$$\Sigma = U\Lambda U^T$$

Then one can rewrite the probability density of $\mathbf{\Theta}$ in the new reference:

$$\begin{aligned} p(\mathbf{\Theta}, \{\bar{\mathbf{\Theta}}, \Sigma\}) &= K(\mathbf{\Theta} - \bar{\mathbf{\Theta}}, \Sigma) \\ &= K(U^T(\mathbf{\Theta} - \bar{\mathbf{\Theta}}), U^T\Sigma U) \\ &= K(U^T(\mathbf{\Theta} - \bar{\mathbf{\Theta}}), \Lambda) \\ &= \Pi_{i=1}^{i=d} K([U^T(\mathbf{\Theta} - \bar{\mathbf{\Theta}})]_i, \lambda_i), \end{aligned}$$

with K being the density function of a Gaussian random variable:

$$K(\mathbf{\Theta}, \Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|}}\, e^{-(1/2)\mathbf{\Theta}\Sigma^{-1}\mathbf{\Theta}}.$$

the last line of the above computation therefore makes use of the diagonal covariance matrix, to decompose the multivariate Gaussian into a product of one-dimensional Gaussians. This approach finds a linear transformation U transforming the data into a space where every components are uncorrelated. Assuming that the Gaussian assumption is true, this transformation is also equivalent to the statistical independence of the components. Assume that the diagonal covariance coefficients are sorted $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq 0$. The terms of the above equation that correspond to the smallest eigenvalues (say λ_i for $i > d_0$) are the directions along which the density is narrowly spread, so that $\mathbf{\Theta}$ almost certainly verify $[U(\mathbf{\Theta} - \bar{\mathbf{\Theta}})]_i \approx 0$. Reducing the dimension of the Gaussian by setting the λ_i to zero for $i > d_0$ is known as *Principal Component Analysis* [89]. As a consequence, this method recovers the components that best explain the data in terms of covariance. The data vector can then be approximated by:

$$\mathbf{\Theta} = \bar{\mathbf{\Theta}} + \sum_{i=0}^{i=d_0} x_i U_i$$

This dimensionality reduction technique also has a probabilistic interpretation [195]. In this approach, the data (observed samples) are reconstructed from the linear transformation of a uniform and unit variance d_0-dimensional Gaussian random variable with an additive Gaussian noise.

Despite the nice theoretical properties of Gaussian densities and the existence of convenient tools to recover their parameters, this model often fails to approximate the samples. This is mostly due to the fact that data distribution forms clusters and each cluster has its own Gaussian behavior. For example, if we consider the deformation of the left ventricle one should expect that the heart surface properties form several clusters, related with age, gender, weight, or pathologies affecting the anatomy.

3.2.3 Gaussian Mixture Models

Gaussian mixture models are examples of multi-modal densities which make the assumption that each cluster can be determined from a normal density. Furthermore, we assume that probability of observing a sample, depends on the prior probability of the cluster and the probability of this sample in this cluster. The general form of the distribution for a mixture of M Gaussians is:

$$p(\mathbf{\Theta}, \mathcal{G}) = \sum_{i=1}^{i=M} \pi_i . K(\mathbf{\Theta} - \mu_i, \Sigma_i), \text{ subject to the constraint } \sum_{i=1}^{M} \pi_i = 1$$

with $\mathcal{G} = \{\mu_i, \Sigma_i, \pi_i\}_{i=1}^{M}$ defining all parameters of the mixture model, π_i being the a priori probabilities for the clusters and $\{\mu_i, \Sigma_i\}$ the mean and variance of the Gaussian density of each cluster. There exists two principal methods to estimate the parameters: K-means and Expectation-Maximization that will be briefly reviewed.

Under the assumption that data are produced from a mixture of Gaussians, the basic idea of these algorithms is to classify the data, (associate it to a source : a single Gaussian) and compute the parameter of each Gaussian with the maximum likelihood method presented in the previous subsection. Mixture of Gaussians estimation does not admit a closed form solution, therefore the proposed methods are iterative and require an initial guess.

K-means [105, 63] is using the most straightforward approach. Starting from a 'guessed' position of the centers $\{\mu_i\}$ for each Gaussian component, every data point is associated with the component whose center is the closest. Once the M clusters of data points are produced, the new position of the cluster center is computed using the empirical mean value. Then one can also determine the other parameters: $\{\Sigma_i\}$ using the assumption that the assigned samples are drawn from a single Gaussian random variable, and $\{\pi_i\}$ by weighing the components according to the number of samples in each cluster.

Fuzzy K-means [63] is an extension of the previous approach, where classification of data

points in clusters is not binary. Also the semantic associated to this method is fully deterministic, we will present Fuzzy K-means in a probabilistic approach as an introduction to its generalization with Expectation-Maximization.

We define the probability that an observation $\mathbf{\Theta}_i$ belongs to the Gaussian component j with center μ_j as a function that depends on the Euclidean distance between the observation and μ_j:

$$m_{ij} = \frac{d(\mathbf{\Theta}_i, \mu_j)^{-2/(\phi-1)}}{\sum_{j=1}^{M} d(\mathbf{\Theta}_i, \mu_j)^{-2/(\phi-1)}},$$

where $\phi > 1$. One notices that this formula naturally verifies $\sum_{j=1}^{M} m_{ij} = 1$. The position of the centers are updated using the weighted sum of $\mathbf{\Theta}_i$ with the probability m_{ij}:

$$\mu_j = (\sum_{i=1}^{N} m_{ij}\mathbf{\Theta}_i)$$

Computed weights ($m_{i,j}$) only depend on the Euclidean distance from the cluster center to the sample. As a consequence, there exists an underlying assumption that all clusters are equivalent and sampled from a mixture of isotropic Gaussian probability. Let us now interpret this assumption considering a density model with the form:

$$p(\mathbf{\Theta}|\mathcal{G}) = \frac{1}{M}\sum_{i=1}^{i=M} K(\mathbf{\Theta} - \mu_i, \sigma.I)$$

with all components having the same weight and diagonal variance. $\mathcal{G}$ is the model, containing in that case the position of the centers and the variance σ. This probabilistic approach to Fuzzy K-means computes the weight m_{ij} as the probability for a point $\mathbf{\Theta}_i$ belonging to the Gaussian component X_j with center μ_j. We write $Class(\mathbf{\Theta})$ the index of the Gaussian component into which a data point $\mathbf{\Theta}$ is falling. Using Bayes rule:

$$\begin{aligned} m_{ij} &= p(Class(\mathbf{\Theta}_i) = j|\mathcal{G}, \mathbf{\Theta}_i) \\ &= \frac{p(\mathbf{\Theta}_i|Class(\mathbf{\Theta}_i) = j, \mathcal{G})p(Class(\mathbf{\Theta}_i) = j)}{p(\mathbf{\Theta}_i|\mathcal{G})} \\ &= \frac{e^{-d(\mathbf{\Theta}_i,\mu_j)^2/\sigma}}{\sum_{j=1}^{M} e^{-d(\mathbf{\Theta}_i,\mu_j)^2/\sigma}}, \end{aligned}$$

where $p(\mathbf{\Theta}_i|Class(\mathbf{\Theta}_i) = j, \mathcal{G})$ is a single component Gaussian density and $p(Class(\mathbf{\Theta}_i) = j) = 1/M$. This expression shows the same 'fuzziness' properties as the historical one, with convergence towards the traditional K-means method when the variance σ goes to 0. K-means is a simple

efficient tool for clustering with known strength and certain limitations. The most critical part is the selection of the number of clusters, and the fact that isotropic densities of the components are considered which often fail to model the actual multimodal distribution.

One can overcome this limitation through a simultaneous estimation of the centers, variance and a priori weights for all components of the mixture. This is performed using an Expectation-Maximization (EM) method where the E-step estimates the class probability of every data point with Bayes rule:

$$\begin{aligned} m_{ij} &= p(Class(\mathbf{\Theta}_i) = j|\mathcal{G}, \mathbf{\Theta}_i) \\ &= \frac{\pi_j K(\mathbf{\Theta}_i - \mu_j, \Sigma_j)}{\sum_{j=1}^{M} \pi_j K(\mathbf{\Theta}_i - \mu_j, \Sigma_j)} \end{aligned}$$

The M-step consists in the maximization of the expected value of $\log[p(\mathbf{\Theta}, Class(\mathbf{\Theta})|\mathcal{G})]$ using the previously determined class probability $p(Class(\mathbf{\Theta}_i) = j|\mathcal{G}', \mathbf{\Theta}_i)$ and considering fixed mixture parameters $\mathcal{G}'$. The conditional expectation is formally written as:

$$Q(\mathcal{G}, \mathcal{G}') = E(\log(p(\{\mathbf{\Theta}\}, \{Class(\mathbf{\Theta})\}|\mathcal{G}))|\mathcal{G}, \{\mathbf{\Theta}'\}),$$

the M-step is the maximization of this criterion with respect to the mixture parameters $\mathcal{G}$:

$$\hat{\mathcal{G}} = \arg\max_{\mathcal{G}} Q(\mathcal{G}, \mathcal{G}')$$

Derivation of Q with respect to the parameters $\mathcal{G}$ leads to the following results for the update of mixture model parameters:

$$\begin{aligned} \pi_j &= \frac{1}{N} \sum_{i=1}^{N} p(Class(\mathbf{\Theta}_i) = j|\mathcal{G}', \mathbf{\Theta}_i) \\ \mu_j &= \frac{\sum_{i=1}^{N} p(Class(\mathbf{\Theta}_i) = j|\mathcal{G}', \mathbf{\Theta}_i)\mathbf{\Theta}_i}{\sum_{i=1}^{N} p(Class(\mathbf{\Theta}_i) = j|\mathcal{G}', \mathbf{\Theta}_i)} \\ \Sigma_j &= \frac{\sum_{i=1}^{N} p(Class(\mathbf{\Theta}_i) = j|\mathcal{G}', \mathbf{\Theta}_i)(\mathbf{\Theta}_i - \mu_j)(\mathbf{\Theta}_i - \mu_j)^T}{\sum_{i=1}^{N} p(Class(\mathbf{\Theta}_i) = j|\mathcal{G}', \mathbf{\Theta}_i)} \end{aligned}$$

It can be proven that one iteration of the EM algorithm always increases likelihood of the samples and therefore the method does converge to a local maximum [116]. This approach is also more robust than K-means and Fuzzy K-means as the underlying model is a better approximation of the true distribution.

However, the performance of EM also depends on the model complexity (the number of Gaus-

sians) and the initialization of the mixture parameters. For example the use of more and more complex models will produce better and better approximations but not necessarily capture the mixture clusters. At the same time the estimation of the model parameters will become more critical due to the increased number of degrees of freedom for the underlying model. Therefore, one should pay attention to the number of mixture components and appropriate effort is to be made in this direction.

The EM-solution to the problem of parameter estimation for the mixture model assumes that the number of Gaussians in the mixture is known a-priori. In order to determine the number of components, two criterion are frequently used:

- The **Schwartz Criterion**[69] is one of the Bayesian information criterion for model selection. In this case, model selection means selecting the number M of Gaussians.

$$J(\mathcal{G}, M) = -\log(p(\mathcal{H}|\mathcal{G})) + M.\log(N)$$

- The **Minimum Description Length**[65] aims to penalize the number of degrees of freedom related to the model according to the assumption that the more parameters the model involves the more expensive it will be to encode. The approximation quality can be combined with a term that penalizes the cost of coding the model parameters. In the case of Gaussian mixtures, a criterion based on length description [147] was added so that the final number M of Gaussian retained in the mixture is minimizing the global criterion:

$$\begin{aligned} J(\mathcal{G}, M) &= -\log(p(\mathcal{H}|\mathcal{G})) + MDL(M) && (3.1) \\ &= -\log(p(\mathcal{H}|\mathcal{G})) + \frac{M}{2}\left[1 + d + \frac{d(d+1)}{2}\right]\log(N) && (3.2) \end{aligned}$$

MDL and Schwartz criteria do not depend on the parameters of $\mathcal{G}$ and have a fixed cost according to the number of mixture models. Therefore, in order to determine the optimal set of parameters that minimizes $J(\mathcal{G}, M)$, M_{max} successive minimization procedures using EM algorithm need to be performed before selecting the optimal model parameters. Although this techniques finds better model fitting, it is M_{max} times more expensive.

Mixture models can go beyond Gaussian densities and be defined using more complex distributions. One can claim that this model is a good compromise between computational complexity and approximation performance. However, in the case where the number of components is reasonable, they make an explicit assumption on the parametric form of the density. In real applications such assumptions are not known or cannot be made in advance. Non parametric densities are an alter-

native to parametric models and aim to describe a model explaining the unknown random variable using all observed samples.

3.2.4 Non Parametric Density Estimation

Non parametric density estimation was initially proposed as a generalization of histograms. Histograms considers a tessellation of the whole space and associate to each tile $'i'$ the value $p(i) = \#\{data\ points\ in\ tile'i'\}/N$. Non parametric density estimation will overcome the constraint on the choice of the tessellation. In such a context, for every observation $\mathbf{\Theta}$ a d-dimensional hypercube $\mathcal{R}(\mathbf{\Theta})$ centered at $\mathbf{\Theta}$ with volume V is considered, and the corresponding density is defined as:

$$\hat{f}(\mathbf{\Theta}) = \frac{\#\{data\ points\ in\ \mathcal{R}(\mathbf{\Theta})\}}{N.V}$$

This function satisfies the property of a density, and $(\hat{f}(\mathbf{\Theta}).V)$ is the estimated probability for a test sample to lie in the hypercube $\mathcal{R}(\mathbf{\Theta})$. If we consider the indicator function of the hypercube $\mathcal{R}(\mathbf{\Theta})$:

$$\mathbb{1}_{\mathcal{R}(\mathbf{\Theta})}(\mathbf{x}) = \begin{cases} 1 & if\ \mathbf{x} \in \mathcal{R}(\mathbf{\Theta}) \\ 0 & else \end{cases} \tag{3.3}$$

One can rewrite the density using the Parzen window estimator [137]:

$$\hat{f}(\mathbf{\Theta}) = \frac{1}{N.V} \sum_{i=1}^{N} 1_{\mathcal{R}(\mathbf{\Theta}_i)}(\mathbf{\Theta})$$

Parzen windows are very popular density approximation techniques with applications to segmentation, registration, tracking, etc. Kernel density estimation refers to a generalization of the non-parametric approximation using an arbitrary interpolation/window function (certain constraints are to be satisfied). Let us consider that the samples $\{\mathbf{\Theta}_i\}_{i=1}^{N}$ are drawn from an unknown random variable taking values in $\mathbb{R}^d$ with density function f. A non parametric estimator will make use of a particular symmetric density function, named kernel, which will also be denoted by $\mathcal{K}$ and

defined in $\mathbb{R}^d$ present the same properties as the centered Gaussian probability density function:

$$\int_{\mathbb{R}^d} \mathcal{K}(\mathbf{\Theta}) d\mathbf{\Theta} = 1$$
$$\int_{\mathbb{R}^d} \mathbf{\Theta}\mathcal{K}(\mathbf{\Theta}) d\mathbf{\Theta} = 0$$
$$\int_{\mathbb{R}^d} \mathbf{\Theta}.\mathbf{\Theta}^T\mathcal{K}(\mathbf{\Theta}) d\mathbf{\Theta} = I_d$$

In its general form, the iso-contours of $\mathcal{K}$ are centered and symmetric with respect to every direction. Then an estimator of the unknown density function can be written in a general form, using a non-singular matrix H:

$$\hat{f}(\mathbf{\Theta}) = \frac{1}{N \det(H)} \sum_{i=1}^{N} \mathcal{K}(H^{-1}(\mathbf{\Theta} - \mathbf{\Theta}_i)).$$

In practice H is directly related to the covariance matrix of the distribution of the kernel. It is therefore more practical to consider this covariance matrix Σ also referred as *bandwidth* which verifies $\Sigma = H^T.H$:

$$\begin{aligned} \hat{f}(\mathbf{\Theta}) &= \frac{1}{N\sqrt{\det(\Sigma)}} \sum_{i=1}^{N} K(\Sigma^{-1/2}(\mathbf{\Theta} - \mathbf{\Theta}_i)) \\ &= \frac{1}{N} \sum_{i=1}^{N} K_\Sigma(\mathbf{\Theta} - \mathbf{\Theta}_i). \end{aligned}$$

The selection of the kernel as well as of the corresponding bandwidth is quite critical in particular when observations are defined on spaces of high dimension. There exists various forms of kernel functions. One can classify them in the following distinct classes [205]:

- *Uniform kernels* with $\mathbb{1}(\|u\| < 1)$ being the kernel function. These kernels correspond to the Parzen window and have interesting practical properties when considering convolution.

- *Epanechnikov kernels* have the form $\frac{3}{4}(1 - u^2).\mathbb{1}(\|u\| < 1)$. This kernel function offers nice theoretical properties but limited practical applicability.

- *Gaussian kernels* have the form $\frac{1}{(2\pi)^{d/2}} \exp(-\|u\|^2/2)$. This form proves interesting especially thanks to its nice behavior under convolution product.

Independently from the selection of the kernel function, the selection of the corresponding bandwidth is equally important. It is non-natural to treat all the samples (with different rate of appearance) in an equivalent fashion. Indeed, a fixed bandwidth approach often produces an undersmoothing in areas with sparse observations and oversmoothing in the opposite case. In order to demonstrate that, let us shortly discuss the expected approximation properties of the kernel estimation. The expected mean of $\hat{f}$ at a given location with respect to the samples is the convolution of the true distribution with the kernel:

$$\bar{f}(\mathbf{\Theta}) = E(\hat{f}(\mathbf{\Theta})) = \int K_\Sigma(\mathbf{\Theta} - \mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

Given the form of the kernel this produces a smooth version of the actual density. Using a second order Taylor expansion of f at $\mathbf{\Theta}$, one can explicitly determine the local bias of the estimator [205]:

$$Bias(\hat{f}(\mathbf{\Theta})) = E(\hat{f}(\mathbf{\Theta})) - f(\mathbf{\Theta}) = \frac{1}{4}\mathrm{trace}[\Sigma.\mathcal{H}_f(\mathbf{\Theta})],$$

where $\mathcal{H}_f$ is the Hessian matrix of the true density function at $\mathbf{\Theta}$. Therefore a kernel density estimator is always biased, and the decrease of the spectral radius of Σ will lead to smaller and smaller bias. However the number of samples in real cases is limited and therefore the choice of the actual bandwidth Σ is critical and not directly related to the statistical bias.

In order to assess the accuracy of the estimator, one may compute the Mean Square Error, with respect to the samples: $MSE(\hat{f}(\mathbf{\Theta})) = E\{(\hat{f}(\mathbf{\Theta}) - f(\mathbf{\Theta}))^2\}$ which produces a local expression of the error depending on the number of samples. In order to assess directly the value of Σ a natural criterion named Asymptotic Mean Integrated Square Error $(AMISE)$ was considered [205]:

$$\begin{aligned} AMISE(\hat{f}) &= \lim_{N\to\infty} \int_{\mathbb{R}^d} MSE(\hat{f}(\mathbf{\Theta}))d\mathbf{\Theta} \\ &= Var\{\hat{f}(\mathbf{\Theta})\} + Bias^2\{\hat{f}(\mathbf{\Theta})\} \\ &\sim \frac{1}{4}\int_{\mathbb{R}^d} \mathrm{trace}[\Sigma.\mathcal{H}_f(\mathbf{\Theta})]^2 d\mathbf{\Theta} + \frac{1}{N\sqrt{(\det(\Sigma))}}\|K\|_2^2, \end{aligned}$$

with $\|K\|_2$ being the L^2 norm of the kernel. The formula will also depend on the Hessian of the density function. The density f being unknown, minimization of such criterion cannot be used directly to estimate the value of the kernel bandwidth. A set of methods to address this issue makes use of an initial assumption on the form of the density, compute the *AMISE* criterion and set the bandwidth of the kernel Σ in order to minimize this quantity (known as Plug-in methods). Such method assume a parametric form for the distribution and retrieve the optimal parameters using maximum likelihood method. The initial estimate of the density will be used to compute $\|K\|_2$

and $\mathcal{H}_f$ in order to assess the size of the kernel with AMISE criterion. Scott's method [161] is one of them where a normal distribution for the whole data is assumed. One estimates its covariance Σ_0 given the data and accesses the Hessian of the density using the obtained density. Based on the closed form solution obtained in the univariate case, this technique will give the general form for the kernel bandwidth: $\Sigma = N^{-2/(d-4)}\Sigma_0$. We have previously introduced 3 different types of kernel functions. One may notice that the Epanechnikov kernel has the property of minimizing the *AMISE* convergence criterion [113]. However, the *AMISE* criterion is asymptotic and it can be proven that using a Gaussian mixture instead does not affect the quality of estimation.

Now that the state of the art on modeling random variables has been presented, let us focus on the problem of modeling shape variations. Medical image analysis and in particular 3D processing and parametric surfaces refer to high-dimensional random variables. Furthermore, organs of the human body exhibit different degrees of variability among individuals. Therefore, a need exists to introduce models of varying complexity where the aim is to produce a reliable density approximation with the smaller possible number of parameters. Such approximation can either be parametric or non-parametric according to the organ complexity.

3.3 Statistical Modeling of Shape Variations

Modeling shape variations is equivalent to studying the deformations between the reference shape model and all samples of the training set. Such a process involves two critical points: (i) the selection of the reference shape, and (ii) the selection of the form of the density that models the deformations towards explaining the training set as optimally as possible.

One should note though that these deformations are due to organ-specific and acquisition-specific reasons. For example, the position of the subject in the scanner may vary from one acquisition to another leading to different spatial positions of the shape samples and eventually affects the deformation model. On top of that, different subjects have different global anatomical properties, and if we consider the heart, one should expect different cardiac volumes. However, in order to study the heart, what is interesting is the local displacement of the myocardium. Therefore, in order to inherit such invariance to the model a global registration is to be performed for all samples where translation, rotation and scale have to be determined. The outcome of this task can then be used to produce a training set where studying the deformations between the reference shape and the training examples is meaningful.

3.3.1 Initial Reference Shape

One of the critical parts of model building is related to the selection of the reference. This model needs to be general enough and at the same time should encode local deformations so that the reference shape can be mapped to any sample in the training set. Towards reducing the complexity of the modeling phase, the degrees of freedom of this deformation model should be as low as possible. The initial guess will be built from a sample of the training set. The element presenting the most important regularity with all the visible properties of the shape of interest will be extracted. In order to illustrate this process let us consider the case of the left ventricle.

3D left-ventricle data were obtained using a semi-automatic segmentation on CT-scans [74]. This algorithm is purely data-based, and make individual assignments at the voxel level. Therefore, segmentation can be erroneous. We manually chose a regular sample of the training set, where the papillary muscles and separation between mitral and aortic valves being well segmented. Furthermore, we impose the constraint of a model with an approximately constant thickness of the myocardium, an expected property of healthy subject. Registration with respect to control points will also address locally the thickness of the muscle, so that the deformed model adequately matches pathological hearts with thinner myocardium in particular areas.

The considered sample shape is smoothed using a Gaussian filter, a signed distance transform is applied and a mesh is generated using the mesher contained in the CGAL library [1]. Figure 3.1 shows successive steps of the reference shape model building. The mesh is manually edited to remove the remaining irregularities of the surface that are specific to the subject.

In the case of 2D shape modeling, we applied our method to sagittal view of the corpus callosum manually segmented from MRI images. The shape being more regular, the choice of an initial reference shape is less critical. We picked one sample of the training set presenting sufficiently regular properties. Post-processing the extracted sample was not necessary in this case.

We should mention though that such a process is only used to provide an initial guess of the reference shape and not the final model. This reference shape will be used to bring all samples in the same pose. Once such a process has been completed, a more accurate reference shape can be obtained. Within this process, invariance to global transformation and pose subtraction is required.

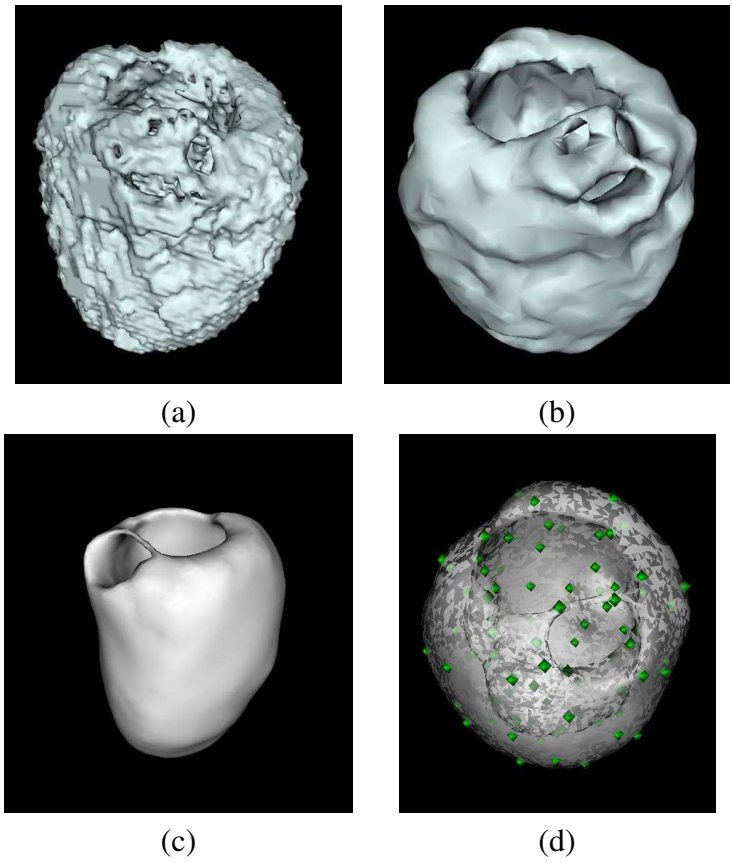


Fig. 3.1: Successive steps of the reference shape model building. (a) Mesh extracted from a automatically segmented ventricle with marching cubes, (b) Smoothed and decimated mesh, (c) Manual manipulation of the model, (d) the reference shape model with the set of control points describing the deformation

3.3.2 Pose Subtraction and Local Registration

Let $\mathcal{C}_{\mathcal{M}}$ be the selected and processed shape resulting from the previous section. This will be our initial reference. This reference will be registered to all the N shapes of the training set named $\{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_N\}$. In most medical application, the pose of an object of interest is due to the position of the patients during the exam. This information is not relevant for the modeling of the anatomy and should be removed from the training set. In the corpus callosum case, an initial shape registration is performed with respect to an affine transformation. Regarding the heart ventricle a similarity transform was used to retrieve the pose information. This is performed for all the N shapes. The inverse of the transformation is applied to every sample, bringing the whole training set to the pose of the reference shape model.

Going to local registration, we have focused in the previous chapter on two classes of spatial

deformations. We have built a coarse to fine registration framework for both of them. Free Form Deformation proposes a set of regular grids with decreasing spacing between control points (see section 2.3.2)). Regarding Thin Plate Splines, control points location was built manually considering the local registration error on the different samples of the training set as proposed in section (2.3.2).

The final parameter vector associated with the registration at the finest scale is retained and used for statistical shape modeling. Uncertainty in the parameter vector is also computed for every registered sample (see section (2.5)). Consequently, every shape $\mathcal{S}_i$ in the training set is associated with a registration vector $\mathbf{\Theta}_i$ and an uncertainty matrix $\Sigma_{\mathbf{\Theta}_i}$.

However, one should recall that this pose subtraction was done towards a reference model initially extracted from the training set. This selection introduces certain bias in the process and the use of uncertainty at that point will be useful, since:

- The registration cannot be perfect due to the limited number of deformation parameters, therefore errors will be naturally related to the reference model.
- Segmentation errors exist in the training set itself.

The initial shape selection does not guarantee that the most representative shape model was selected. This constraint is important since it will propagate to the model the ability to explain deformations from the reference to all samples using a small number of degrees of freedom. Therefore the initial shape model will be transformed to give a more accurate representation of all the samples.

3.3.3 Average Reference Shape Model

The idea of building an average shape model actually find its foundation in Active Shape Models [46] where the statistical model is parametric and assumes a Gaussian distribution. Using a reference shape that is a true mean of the training set makes the covariance estimator unbiased. However the justification in terms of statistical bias is questionable given the limited size of the training set. The actual justification is practical: it was previously stated that the registration framework is heavily dependent on the initialization. Choosing a reference shape capable of being deformed evenly to every shape of the training set should be favored. This question should be raised with an appropriate local metric defined in the shape manifold [138, 96]. Our approach is different as it is based on deformations: a shape of the training can be explained using different

deformation parameters. This explains the sensitivity of the registration with respect to the initial state. Also the matrix of uncertainty (see section (2.5)) gives a local estimate of the deformation candidates representing similar shapes. We have used the Mahalanobis distance with the estimated uncertainty for each sample to compute the mean shape over the training set. Updating the model is performed iteratively, in the same fashion as in Active Shape Models [46]:

1. Register the whole training set using $\mathcal{C}_{\mathcal{M}}$ as common source shape, compute uncertainties $(\{\mathbf{\Theta}_i, \Sigma_{\mathbf{\Theta}_i}\})$.

2. Compute the mean parameter vector:

$$\bar{\mathbf{\Theta}} = \arg\min_{\mathbf{\Theta}}(\sum_i (\mathbf{\Theta} - \mathbf{\Theta}_i)^T \Sigma_{\mathbf{\Theta}_i}^{-1} (\mathbf{\Theta} - \mathbf{\Theta}_i)).$$

3. Update the reference shape: $\mathcal{S}_{\mathcal{M}} \leftarrow \mathcal{L}(\mathcal{S}_{\mathcal{M}}, \bar{\mathbf{\Theta}})$.

4. Update the control points initial position.

In practice, a stable mean shape is obtained after a couple of iterations of the above process.

Once the reference shape has been determined, the next step consists of modeling the variations of the deformation estimates from this shape to all examples of the training set. The amount of variation between training examples heavily depends on the organ under consideration. Organs like the heart refer to smooth surfaces with relatively limited variation. In order to account for such degree of variation, we will introduce two models of increasing complexity. In order to do so, we need clinical data with experts/physicians segmentation. However, the data obtained from clinical experts in many cases exhibit important variations between them, and these errors will be propagated to the model. Furthermore, as explained earlier, the registration process will inevitably produce a deformation result with some erroneous measurements. Last, geometric correspondences do not always correspond to anatomical ones, and that is a third source of error propagation in the model.

Therefore a need exists to propose models which somehow account for these errors. In this domain, qualitative interpretation of the results is equally important with the quantitative solution. Therefore, understanding or providing measurements of uncertainty along with the obtained solution is a valuable element to data interpretation. In the next section we propose two models for the observed deformations, being able to deal to some extent with the above issues.

3.4 Dimensionality Reduction with Linear Methods

This section will focus on the use of uncertainty to build a statistical prior on the deformation parameters $\mathbf{\Theta}$. Having the uncertainty defined for every sample, we consider in this section that every shape of the training set contains sufficient information to represent a continuous cluster of possible shapes. This cluster is defined using a Gaussian random variable, centered on the registration result $\mathbf{\Theta}_i$ with the uncertainty $\Sigma_{\mathbf{\Theta}_i}$ as covariance matrix. This section will use the covariance matrices as defined and scaled in section (2.5.5) based on the physical interpretation of the bandwidth. From these elements we will estimate an a priori density function $\hat{f}(\mathbf{\Theta})$ that models the variations of the training set. The simplest possible model to be considered is the one assuming that the variables follow a Gaussian density. In such a context, principal components analysis can be used to describe the observed density of the deformations.

3.4.1 Uncertainty-Driven Principal Component Analysis with Uncertainty.

Assuming a Gaussian distribution of the whole training set, we first apply the Principal Component Analysis technique on the data. In the present case however the data is not simply a set of vectors $\{\mathbf{\Theta}_i\}$, but a continuous distribution of vectors with probability density:

$$\hat{f}(\mathbf{\Theta}) = \frac{1}{n}\sum_i K(\mathbf{\Theta} - \mathbf{\Theta}_i, \Sigma_i)$$

The maximum likelihood estimator of the mean and variance can be computed as in section (3.2.2):

$$\begin{aligned}
\hat{\mu} &= \int (\mathbf{\Theta} - \mathbf{\Theta}_i) f(\mathbf{\Theta}) d\mathbf{\Theta} \\
&= \frac{1}{n}\sum_{k=1}^{N} \mathbf{\Theta}_k \\
\hat{\Sigma} &= \int (\mathbf{\Theta} - \hat{\mu}).(\mathbf{\Theta} - \hat{\mu})^t f(\mathbf{\Theta}) d\mathbf{\Theta} \\
&= \frac{1}{n}\sum_i \int \left((\mathbf{\Theta} - \mathbf{\Theta}_i) + (\mathbf{\Theta}_i - \hat{\mu})\right) . \left((\mathbf{\Theta} - \mathbf{\Theta}_i) + (\mathbf{\Theta}_i - \hat{\mu})\right)^t K(\mathbf{\Theta} - \mathbf{\Theta}_i, \Sigma_i) d\mathbf{\Theta} \\
&= \frac{1}{n}(\mathbf{\Theta}_i - \hat{\mu}).(\mathbf{\Theta}_i - \hat{\mu})^T + \sum_{k=1}^{N} \Sigma_i
\end{aligned}$$

The estimated mean is the same if we do not consider the uncertainty. The global covariance matrix sums up the uncertainty associated with every sample and the sample covariance of the training set

(covariance between classes). All covariances being added, uncertainty brings additional noise to the standard Principal Component Analysis. This is visible as the decrease of the eigenvalues is much slower when considering uncertainties and using PCA would require considering a greater number of principal directions. Therefore such a model does not benefit much from the introduction of uncertainties and more advanced dimensionality reduction techniques and linear models are to be determined.

3.4.2 Uncertainty-Driven Independent Component Analysis.

Independent Component Analysis (ICA) also known as Blind Source Separation is a statistical and computational technique that can determine the statistical nature of sets of random variables, measurements, or signals. This technique is based on a linear decomposition of a multidimensional model, just as Principal Component Analysis (that computes directions of largest variance) and Factor Analysis (that explains statistical measurements as a linear combination of a certain number of factors).

While PCA retrieves an orthogonal transformation of the data into uncorrelated components, ICA linearly transforms the data into components that are statistically independent. This means that the projections of the observed data along these directions form a set of independent scalar random variables [86]. The approach proposed with factor analysis uses a set of candidate factors to retrieve the weights that generate the one dimensional observed data, while ICA directly uses multivariate data to estimate both the factor and the weights.

In brief ICA relies on a generative model that assumes that data can be generated as a linear combination of random variables. These latent variables are assumed non-Gaussian and mutually independent and are called the *independent components* of the observed data. Regarding our application to modeling of shape deformations, this means that any sample shape can be explained as a linear combination of specific independent deformations of the reference shape. The simplest proof of independence is found in the fact that it affects different parts of the shape under study:

$$\mathbf{\Theta} = \mathbf{A}.\mathbf{s}, \tag{3.4}$$

where $\mathbf{\Theta}$ is the observed data vector, here considered as a random variable, $\mathbf{s}$ an m-dimensional random vector of sources with independent components, and $\mathbf{A}$ a constant *mixing* matrix. The ICA process consists in recovering both the source and the mixing matrix using only the observed data. Several approaches have been developed to robustly compute statistical independence of random variables and estimate the values of the inverse *de-mixing* matrix for a fixed number of independent

components:

$$\hat{\mathbf{s}} = \mathbf{W}.\mathbf{\Theta} \approx \mathbf{A}^{+}.\mathbf{\Theta}, \tag{3.5}$$

where $\mathbf{A}^{+}$ refers to the pseudo-inverse of the mixing matrix. Due to the limited number of samples and the choice of the method, there is no equality between these two terms.

Various approaches to estimate the independent sources were presented like InfoMax [9], Jade [23] or Fast-ICA [87, 86]. These methods involve a pre-processing of the data with aim to reduce the dimensionality of the problem which involves two steps:

- *centering:* the mean value of the training set is subtracted from every sample: $(\mathbf{\Theta}_i \leftarrow \mathbf{\Theta}_i - \bar{\mathbf{\Theta}})$. The problem being linear, centering the data is equivalent to centering the independent sources.

- *whitening:* a linear transform is applied to the data so that components become uncorelated with variance 1. This transformation is produced from the Principal Component Analysis applied to the set of data. Obtaining $\tilde{\mathbf{\Theta}}_i = L.\mathbf{\Theta}_i$ with $\sum \tilde{\mathbf{\Theta}}_i \tilde{\mathbf{\Theta}}_i^T = \mathbf{I}$. The utility of whitening resides in the fact that the mixing matrix $\tilde{A}$ for the new data is orthonormal.

  $$E\{\tilde{\mathbf{\Theta}}_i \tilde{\mathbf{\Theta}}_i^T\} = \tilde{A} E\{\mathbf{s}.\mathbf{s}^T\} \tilde{A}^T = \tilde{A}\tilde{A}^T = \mathbf{I}.$$

  Consequently the complexity of the problem is reduced to $d(d-1)/2$, the number of degrees of freedom of an orthonormal matrix. Whitening is also used to perform an initial dimension reduction, by using standard PCA.

The principle difficulty of ICA comes from the definition of statistical independence and the ways to measure it efficiently. In our application, we have focused on the Fast-ICA algorithm [87] that interestingly relates independence and non-Gaussianity. The intuition comes from the Central Limit Theorem, and the fact that any linear combination of independent centered and reduced random variables will have a distribution closer to a Gaussian than any of the two input variables. One will try to maximize the non-Gaussianity of the following expression:

$$y = \mathbf{w}^T.\mathbf{\Theta} = (\mathbf{w}^T.\mathbf{A}).\mathbf{s}.$$

If $(\mathbf{w}^T.\mathbf{A})$ has more than one non-zero coefficients the resulting random variable will be closer to a Gaussian than any of the component of $\mathbf{s}$. Consequently $(\mathbf{w}^T.\mathbf{A})$ has only one non-zero component, and $\mathbf{w}$ corresponds to a line of the de-mixing matrix $\mathbf{W}$.

Various criteria were proposed to measure the non-Gaussianity, examples are the *kurtosis* (sen-

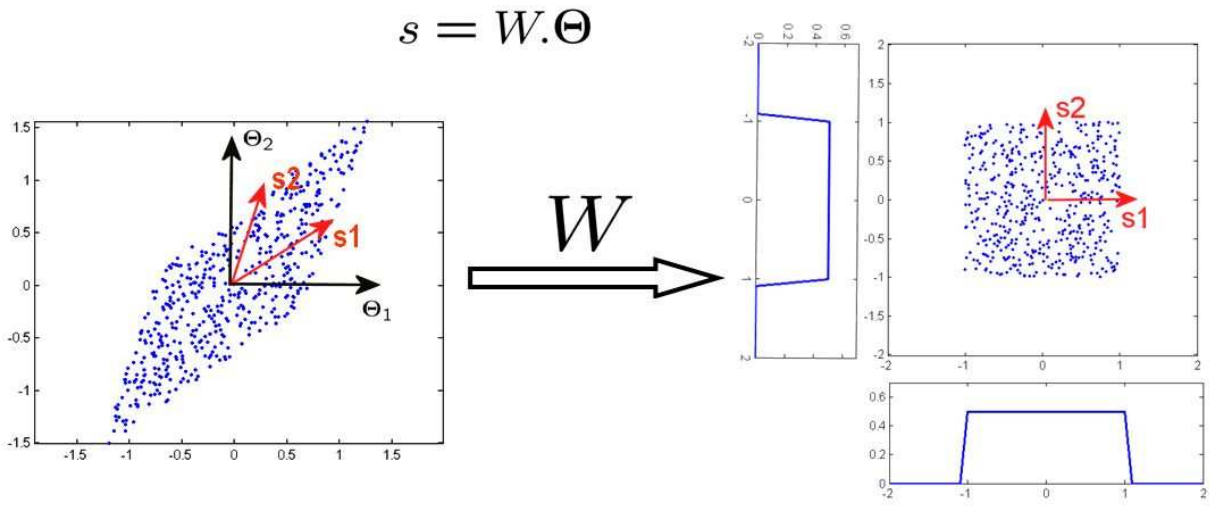


Fig. 3.2: A synthetic example of Independent Component analysis in the 2 dimensional case. ICA will retrieve the linear function that will transform the distribution into a domain where the two components are statistically independent.

sitive to noise), or the *negentropy*. The latter is based on the difference between the estimated entropy of the samples and the entropy of a unit variance Gaussian. An efficient implementation of the algorithm was used in our application [77].

ICA was previously used in shape modeling [198] where the authors proposed a method to order and select the "most relevant" independent components. Their method represents shapes using the point distribution models in a very high dimensional space. Our representation of deformations is in a much lower dimensional space. Thus, our application does not require classification of independent components, we therefore directly compute m components. The components of $\mathbf{s}$ being independent, their distributions can be expressed as the product of 1-dimensional random variables densities:

$$p(\mathbf{s}) = \prod_{i=1}^{m} p_i(s_i) \tag{3.6}$$

where a Gaussian mixture model is fitted with the EM algorithm (section 3.2.3) to estimate the actual density of each component:

$$p_i(s_i) = \sum_{j=1}^{M_i} \alpha_j \mathcal{N}(\mu_{i,j} - s_i, \sigma_{i,j}). \tag{3.7}$$

In practice, the choice of $M_i = 2$ is sufficient to describe the variability of deformations for the samples of the training set. Fig. 3.3 shows the training samples projected on the first three

components. This method can produce a compact statistical model given a training set where all samples are brought to the same pose (section 3.3). However, it does not account for the uncertainties determined during the registration process.

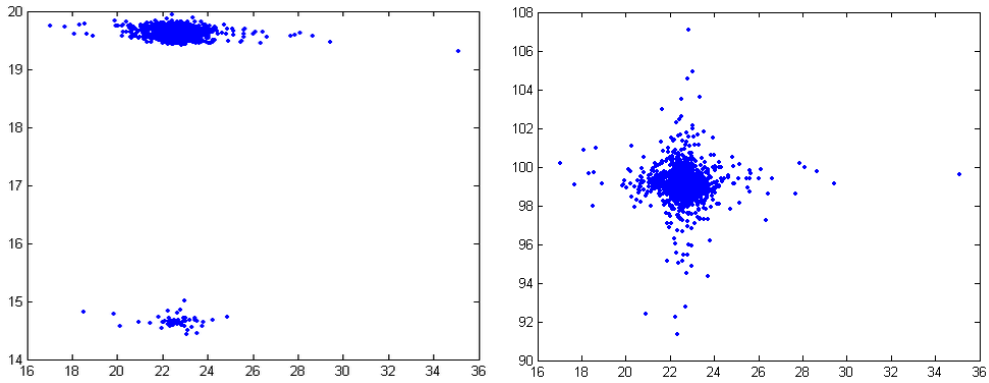


Fig. 3.3: Density plot of independent components $\mathbf{s} = \mathbf{W}.\mathbf{\Theta}$, independence and Gaussian mixture estimation of components appears relevant.

In order to include the registration uncertainties in the process, we generate additional data prior to ICA. Our approach consists of sampling each measurement on the training set using the normal law centered on the registration results with uncertainty being the covariance matrix $\mathcal{N}(\mathbf{\Theta}_i, \Sigma_{\mathbf{\Theta}_i})$. For each example the same number of samples is drawn using the corresponding densities. Such a process will lead to an augmentation of the training set where samples of low uncertainty will have more influence on the model to be recovered. On the other hand, for registration results showing significant values of the covariance matrix determinant, the drawn samples will be dispersed and therefore will have limited influence in the model construction. The augmented set of variables can now be used within the Independent Component Analysis framework to produce a multi-component density of deformations and obtain the following density estimator:

$$\hat{f}(\mathbf{\Theta}) = \Pi_{i=1}^{m} p_i([\mathbf{W}.\mathbf{\Theta}]_i),$$

where $[\mathbf{W}.\mathbf{\Theta}]_i$ is the i^{th} independent component extracted from the data. Such a parametric model is able to propagate the uncertainties and the errors due to segmentation and registration to the model. However, it suffers from an explicit limitation that is related to an explicit assumption being made regarding the behavior of samples. In other words this model will perform well if and only if the observations satisfy these assumptions. If this is not the case the result will be erroneous.

Organs of human body undergo various degrees of variability. Therefore in some cases the

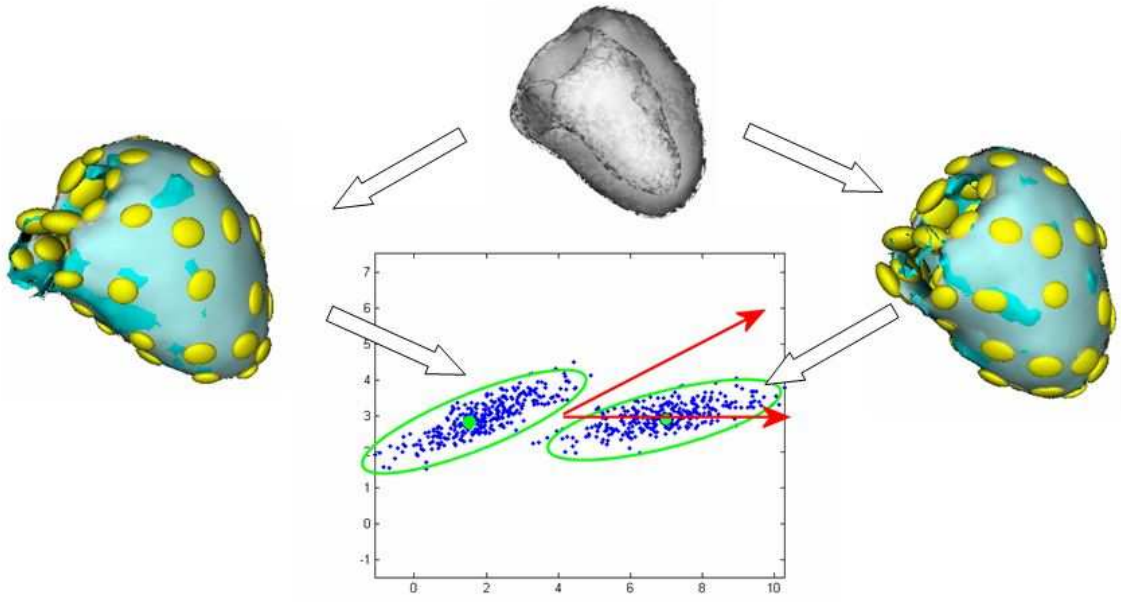


Fig. 3.4: The sampling of deformation candidates according to the uncertainty evaluation, as a prior to Independent Component Analysis on the training set.

use of the above-mentioned model could provide excellent approximation of the density. However in cases where this is not possible, the use of non-parametric models could deal with the non-parametric nature of the samples. In such a context the key will be the propagation of the uncertainties to the non-parametric density.

3.5 Variable Bandwidth Non-Parametric Approximation of Deformations

The use of these estimators in our application is straightforward: Assuming the set of observed shapes is a random variable X admitting a probability density function f, it is desirable to build an estimator for f which we denote $\hat{f}$. We dispose of a random samples of shapes constituting the training set $\{\boldsymbol{\Theta}_i\}_{i=1}^{M}$. These values are not simple realization of the random variables but comes along with uncertainties that characterize the allowable variations of the deformation parameters.

The most basic estimator is the fixed bandwidth kernel density estimator (section 3.2.4) and consists of:

$$\hat{f}(\boldsymbol{\Theta}) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{K}_{\boldsymbol{\Sigma}}\left(\boldsymbol{\Theta} - \boldsymbol{\Theta}_i\right) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\|\boldsymbol{\Sigma}\|^{1/2}} \mathrm{K}\left(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_i)\right) \tag{3.8}$$

where $\mathbf{\Sigma}$ is a symmetric definite positive matrix - often called a bandwidth matrix - that controls the width of the kernel around each sample point $\mathbf{\Theta}_i$.

However, when considering training sets of limited size, usefulness of varying bandwidths is widely acknowledged to estimate long-tails on multi-modal density functions with kernels. In the literature, kernel density estimation methods that do rely on such varying bandwidths are generally referred to as *adaptive kernel* density estimation methods. It adapts to the sparseness of the data by using a broader kernel over observations located in regions of low density. Two useful state of-the-art variable bandwidth kernels consists of the *sample point estimator* and the *balloon estimator* [158, 159].

The *sample point estimator* refers to a covariance matrix depending on the repartition of the points constituting the sample :

$$\hat{f}_S(\mathbf{\Theta}) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\|\mathbf{\Sigma}(\mathbf{\Theta}_i)\|^{1/2}} \mathrm{K}\left(\mathbf{\Sigma}(\mathbf{\Theta}_i)^{-1/2}(\mathbf{\Theta}-\mathbf{\Theta}_i)\right) \tag{3.9}$$

where $\Sigma(\mathbf{\Theta}_i)$ is the smoothing matrix (covariance) associated with $\mathbf{\Theta}_i$. A common selection of $\Sigma(\mathbf{\Theta}_i)$ takes the form:

$$\mathbf{\Sigma}(\mathbf{\Theta}_i) = \sigma(\mathbf{\Theta}_i)\cdot\mathbf{I}. \tag{3.10}$$

An optimal choice of $h(\mathbf{\Theta}_i)$ showing good asymptotic behaviour was proposed as a function of the true distribution ($f(\mathbf{\Theta}_i)^{-1/2}$ in [19] and $f(\mathbf{\Theta}_i)^{-1/d}$ in [2]) and in practice used as a rough estimate of the density based on fixed bandwidth kernel to locally compute the bandwidth factor [78]. One can consider various alternatives to determine the bandwidth function. This method was also useful for energy minimization using mean-shift as presented in [41, 29, 40]. Within our framework, such estimator may be directly used with the uncertainties calculated in section 2.5 and $\mathbf{\Sigma}(\mathbf{x}_i) = \mu\Sigma_{\Theta_i}$.

There is a different approach of the result based on the origin of the uncertainties. These were designed as the possible variations of deformations representing a shape of the training set. Consequently the uncertainty at $\mathbf{\Theta}_i$ defines locally a metric in the space of the deformations. If samples with uncertainty were defined anywhere in the space, one would consider a Riemannian geodesic to compute the distance between two samples. As only a few samples are available, the distance from a vector $\mathbf{\Theta}$ to the shape with parameter $(\mathbf{\Theta}_i, \Sigma_{\mathbf{\Theta}_i})$ will be defined with the Mahalanobis distance: $(\mathbf{\Theta}_i - \mathbf{\Theta})\Sigma^{-1}_{\mathbf{\Theta}_i}(\mathbf{\Theta}_i - \mathbf{\Theta})$. Therefore the sample point estimator can be interpreted as a fixed bandwidth kernel estimator in such a modified vector space by.

$$\hat{f}_S(\mathbf{\Theta}) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\|\sigma.\Sigma_{\mathbf{\Theta}_i})\|^{1/2}} \mathrm{K}\left((\sigma\Sigma_{\mathbf{\Theta}_i})^{-1/2}(\mathbf{\Theta}-\mathbf{\Theta}_i)\right) \tag{3.11}$$

where σ is a global scaling coefficient added for proper smoothing of the estimator.

In the present work we often have an estimation of the uncertainty on the point to be valued. These uncertainties may be produced from shape registration as presented in the previous chapter or based on image information in the case of segmentation. Consequently it is useful to introduce a different kind of estimator with a varying bandwidth at each estimation point. It is known as *balloon estimator* [188]. Bandwidth adapts to the point of estimation depending on the shape of the sampled data according to:

$$\hat{f}_B(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\|\boldsymbol{\Sigma}(\boldsymbol{\Theta})\|^{1/2}} \mathrm{K}\left(\boldsymbol{\Sigma}(\boldsymbol{\Theta})^{-1/2}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_i)\right) \tag{3.12}$$

Considering bandwitdth with the form $\Sigma(\boldsymbol{\Theta}) = h(\boldsymbol{\Theta})\cdot\mathbf{I}$, many estimators where proposed using the distance from a point $\boldsymbol{\Theta}$ to its k^{th} nearest neighbor in the training set. However this kind of choice presents some drawbacks, the variation of the bandwidth is discontinuous and these discontinuities appear directly on the resulting estimator, also the estimator usually fails to integrate to 1. Using the interpretation in terms of metric, given the uncertainty estimate, one will use:

$$\hat{f}_B(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\|\sigma\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}\|^{1/2}} \mathrm{K}\left((\sigma\boldsymbol{\Sigma}_{\boldsymbol{\Theta}})^{-1/2}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_i)\right)$$

Let us consider $\{\boldsymbol{\Theta}_i\}_{i=1}^{N}$ a multi-variate set of measurements where each sample $\boldsymbol{\Theta}_i$ is associated with uncertainty measures in the form of a covariance matrix $\Sigma_{\boldsymbol{\Theta}_i}$. Our objective can be stated as follows: estimate the probability density of a new measurement $\boldsymbol{\Theta}$ that is associated with covariance matrix $\Sigma_{\boldsymbol{\Theta}}$.
Let $\mathbf{X}$ be the random variable characterizing the distribution of deformation of the shape of interest with density f. Practically, f may be estimated with $\hat{f}_S$ using the *sample point estimator*. Therefore $\hat{f}_S$ is expressed in the form $\hat{f}_S = \sum \hat{f}_{S,i}$ where $\hat{f}_{S,i}$ is the density associated with a single kernel $\boldsymbol{\Theta}_i$ with uncertainty $\Sigma_{\boldsymbol{\Theta}_i}$. A new shape sample can be seen as the realization of the same phenomenon. A measure of this phenomenon is observed with a certain amount of uncertainty, modeled as the random variable $\hat{\mathbf{Y}}$ with density $\hat{g}$.

In order to estimate the probability density at the new sample location, one should first determine for all possible $\mathbf{u} \in \mathbb{R}^d$ their contribution to the *distance* from $\hat{f}(\mathbf{u})$ the existing prior estimate of the training set $\hat{\mathbf{X}}$ and weigh them according to their fit with the density function associated to

the test sample $\hat{\mathbf{Y}}$:

$$f_G(\hat{\mathbf{Y}}) = \int \hat{f}_S(\mathbf{u})\hat{g}(\mathbf{u})d\mathbf{u} = \int \left[\sum_{i=1}^{M} \hat{f}_{S,i}(\mathbf{u})\right] \hat{g}(\mathbf{u})d\mathbf{u} = \sum_{i=1}^{M} \left[\int \hat{f}_{S,i}(\mathbf{u})\hat{g}(\mathbf{u})d\mathbf{u}\right] \tag{3.13}$$

In the case of a single Gaussian kernel for the estimates $\hat{g}$ and the $\hat{f}_{S,i}$, We use the interesting convolution properties of the Gaussian function and recover the expression:

$$\hat{f}_G(\mathbf{\Theta}) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\|\sigma^2(\Sigma_{\mathbf{\Theta}} + \Sigma_{\mathbf{\Theta}_i})\|^{1/2}} K((\sigma^2(\Sigma_{\mathbf{\Theta}} + \Sigma_{\mathbf{\Theta}_i}))^{-1/2}(\mathbf{\Theta} - \mathbf{\Theta}_i)) \tag{3.14}$$

where σ is a global scaling coefficient added for proper smoothing of the estimator. We have used the same choice as Cremers *et al.* in [47] using the mean of the minimum distance between sample values, using the Mahalanobis distance associated with the resulting bandwidth $\Sigma_{\mathbf{\Theta}_i} + \Sigma_{\mathbf{\Theta}_j}$:

$$\sigma^2 = \frac{1.5^2}{N^2}\sum_{i} \min_{i \neq j} (\mathbf{\Theta}_i - \mathbf{\Theta}_j)^T (\Sigma_{\mathbf{\Theta}_i} + \Sigma_{\mathbf{\Theta}_j})^{-1} (\mathbf{\Theta}_i - \mathbf{\Theta}_j)$$

The probabilistic interpretation is straightforward: given two measurements of the same phenomenon with two different tools, the random variables $\hat{\mathbf{X}}$ (associated with the sample point estimator) and $\hat{\mathbf{Y}}$ (associated with the new sample), it is desirable to know how likely it would be for these random variables to be equal. To assess this value, one consider the random variable $\hat{\mathbf{Z}} = \hat{\mathbf{X}} - \hat{\mathbf{Y}}$, the previous result is the density function of $\mathbf{Z}$ at point 0:

$$p_{\hat{\mathbf{Z}}}(0) = \int \hat{f}_S(\mathbf{u})\hat{g}(\mathbf{u})d\mathbf{u} \tag{3.15}$$

The present concept could be relaxed to address the case of non Gaussians kernels according to a *hybrid* estimator:

$$\hat{f}_G(\mathbf{\Theta}) = \frac{1}{M}\sum_{i=1}^{M} \frac{1}{\|\sigma(\Sigma_{\mathbf{\Theta}}, \Sigma_{\mathbf{\Theta}_i})\|^{1/2}} \mathrm{K}(\sigma(\Sigma_{\mathbf{\Theta}}, \Sigma_{\mathbf{\Theta}_i})^{-1/2}(\mathbf{\Theta} - \mathbf{\Theta}_i)) \tag{3.16}$$

Such a density estimator takes into account the uncertainty estimates both on the training sample points and on the new observed sample point $\mathbf{\Theta}$. Consequently, the density estimation decreases more slowly in directions of large uncertainties of the sample point, than in any other direction.

This measure can now be used to assess the probability for a new sample of being part of the training set through a process that evaluates the *hybrid* kernel for each of the examples in

the training set. The resulting approach can account for the non-parametric form of the observed density. However, this technique is very time consuming since the computation is linear in the number of samples in the training set, and the computation of a *hybrid* kernel involves a matrix inverse of the sum of covariance matrices. Therefore, there is an eminent need on decreasing the cardinality of the set of retained kernels in the case of large training sets.

3.6 Dimensionality Reduction and Non-Parametric Interpretation of Samples

So as to qualify the fit of the probability density induced by a selection of kernels to the complete training set, we make the selection of kernels that will maximize the log-likelihood of the probability density calculated on the whole training set.

Consider a selection $\mathcal{Z}_K = \{X_1, X_2, \ldots, X_K\}$ of K kernels extracted from the training set. These have associated means and uncertainties $\{\mathbf{\Theta}_i, \Sigma_i\}_{i=1}^{|\mathcal{Z}_K|}$. Then the probability of any registered shape with associated kernel $Y = \{\mathbf{\Theta}, \Sigma\}$ has the form :

$$P_{\mathcal{Z}_K}(Y) = \frac{1}{|Z_K|} \sum_{X_i \in Z_K} K(X_i, Y) \tag{3.17}$$

and $K(X_i, Y)$ corresponds to the calculation of the hybrid kernel estimator:

$$K(X_i, Y) = \frac{1}{\|\sigma(\Sigma_{\mathbf{\Theta}_i} + \Sigma_{\mathbf{\Theta}})\|^{1/2}} K((\sigma(\Sigma_{\mathbf{\Theta}_i} + \Sigma_{\mathbf{\Theta}}))^{-1/2}(\mathbf{\Theta}_i - \mathbf{\Theta}))$$

For a selection of kernels $\mathcal{Z}_K$, one can evaluate the log-likelihood for the entire training set with the associated kernels $\{Y_i\}_{i=1}^N$:

$$C_K = \sum_{i=1}^{N} \log(P_{\mathcal{Z}_K}(Y_i)) \tag{3.18}$$

The best selection of K kernels explaining the whole training set will maximize the above quantity. We use an efficient sub-optimal iterative algorithm to update the set $\mathcal{Z}_K$. A new kernel Y is extracted from the training set as the one maximizing the quantity C_{K+1} associated with $\mathcal{Z}_{K+1} = \mathcal{Z}_K \bigcup Y$. One kernel may be chosen several times in order to preserve an increasing sequence of C_K when adding new kernels. Consequently the selected kernels Y_i used to evaluate the global density probability have prior weight ω_i. We finally use the sample point estimator to assess the

value of a deformation vector $\mathbf{\Theta}$:

$$\hat{f}_S(\mathbf{\Theta}) = \sum_{i=1}^{K} \frac{\omega_i}{\|\sigma\Sigma_{\mathbf{\Theta}_i}\|^{1/2}} K(\sigma\Sigma_{\mathbf{\Theta}_i}^{-1/2}(\mathbf{\Theta} - \mathbf{\Theta}_i)), \text{ with } \sum_{i=1}^{K} \omega_i = 1. \tag{3.19}$$

And the hybrid estimator in the case where a deformation vector $\mathbf{\Theta}$ is available with uncertainty:

$$\hat{f}_G(\mathbf{\Theta}) = \sum_{i=1}^{K} \frac{\omega_i}{\|\sigma(\Sigma_{\mathbf{\Theta}} + \Sigma_{\mathbf{\Theta}_i})\|^{1/2}} K((\sigma\Sigma_{\mathbf{\Theta}} + \Sigma_{\mathbf{\Theta}_i})^{-1/2}(\mathbf{\Theta} - \mathbf{\Theta}_i)), \text{ with } \sum_{i=1}^{K} \omega_i = 1. \tag{3.20}$$

This approach may only be used in the case of a large training set, which contains redundant information. We have applied it to the modeling of corpus callosum, reducing the size of the training set from 80 to 50.

3.7 Conclusion

In this section we have presented a set of methods for modeling shape deformations and proposed different use for the uncertainty information. Using linear parametric methods, we have proposed the use of Independent Component Analysis combined with one dimensional mixture of Gaussians density estimation. This approach produces a reduced representation of the deformation, but also relies on the assumption that the data is produced from the linear combination of independent sources. A different approach that does not make any assumption on the distribution of the data was tested, based on kernel density estimation. Inspired from works on fixed and variable bandwidth kernel methods, we have introduced the uncertainty measure in the estimator, using Gaussian kernels. In the latter case we have proposed a hybrid density estimator capable of accounting for the measure of uncertainty for the shape to be valued.

Having robustly defined the a priori density estimator given the training set, we will present in the next chapter how this information is used to perform segmentation of the left ventricle on CT images.

Future direction following the present work may include the introduction of a hybrid estimator within the Independent Component Analysis method. Also the non-parametric approach may be further explored having in mind the Riemannian interpretation of the space of deformation, and the kernel estimators expressed in this modified metric space.

Chapter 4

Knowledge-Based Segmentation

Abstract – *Segmentation is one of the central problems in the field of computer vision and medical image analysis. It aims to define a partition of the image domain into a set of disjoint regions delineating objects with similar features in an appropriately defined space. Therefore the problem is often equivalent to an optimization problem using edge, intensities, texture and other statistical information of the image. Model free segmentation approaches have considered edge or regional image information. However, in many cases, either due to noise, occlusions or absence of discrimination between the objects and the background, the use of prior knowledge on the object to segment is necessary. Consequently, hard and soft constraints are added to the segmentation results accounting for prior knowledge.*

In this chapter we review the state of the art in model free and model based segmentation and introduce two novel approaches to account for prior knowledge. One is using linear approximation models with uncertainties for the segmentation of the left ventricle. The second assumes complex statistical models based on kernels with variable bandwidth. This model is considered to segment parts of the brain. Both methods inherit the uncertainties of the registration process.

4.1 Introduction

Image segmentation is one of the most fundamental problems in low and mid-level vision. Defining such a problem is not straightforward since it has several variants. One of the most general definition consists of saying that segmentation is equivalent to image partitioning into perceptually meaningful and similar regions according to some criterion. However, the definition of perceptually similar structures is not universal. The term segmentation can be used for simple intensity-based grouping and for object extraction, in images where the notion of similarity in the feature space is not necessarily visually perceivable. In both cases the problem is ill-posed. In the most general setting we have no clue either on the number of objects/regions, or on the statistical characteristics of the regions and features to be extracted.

Image-based similarity techniques can be either intensity/feature-based or knowledge-based. In the first case, the problem becomes equivalent to feature clustering and boundary extraction, while in the second case we are mostly interested in fitting a predefined model to the image. The easiest way to define segmentation consisted in partitioning the image into regions with similar feature properties. Such similarity is defined on the observation space (image/color intensities). Decisions can be taken either at the pixel or at the region level using well studied statistical tests. In order to improve robustness to noise and to avoid individual erroneous decisions, one can impose some spatial consistency constraints on the decision process. Prior works which have considered decisions at the pixel level consist of EM-segmentation [207], clustering using K-means [200], mean-shift [32, 39], etc. On the other hand when aiming to introduce constraints on the decision process such that neighboring pixels are assigned to the same class, one can refer to work on MRFs [12, 216] as well as their recent variants using max-flow/min-cut principles [18], region growing, split-and-merge, etc. These methods are not so related to our approach, therefore will not be developed further and we refer to [170] for their systematic review. These methods can be used beyond conventional intensity-based segmentation through the development of statistical descriptors on any perceptually meaningful space like for example texture [131, 152].

Boundary extraction is an alternative to statistical clustering. Instead of seeking individual decisions for each pixel/voxel, the aim of these methods is to determine image transitions between classes. However the definition of transitions is not trivial since, similar to the perceptual grouping problem, the space on which discontinuities are searched depends heavily on the application. Edge-detection often relies on discontinuities in the visual space. The use of local features and the extraction of local maxima on the image gradient space was the starting point for these methods [22, 79]. The main limitation of these methods was the lack of continuity on the separation

between different regions (decisions were taken again individually). Snakes/active contours were introduced in the mid-eighties and were a step forward along this direction where boundary extraction was equivalent to the minimization of a cost function over a "continuous" curve capturing the desirable image features. Numerous variants based on the original snake formulation were and are still proposed. These aim at improving their capture range as well as dealing with the topology-preserving limitations that were introduced. One can cite their geometric alternative, the geodesic active contour model [24], as well as their implicit, topology-free level set formulation [127]. In order to improve robustness to noise as well as their capture range these methods were amended to use regional information as well either in an explicit [123, 171, 218] or implicit formulation [130, 26]. The proposed framework bears some concept similarities with the original snake approach as well as its level set variant. Therefore related work will be further introduced in the upcoming sections.

The abstract definition of the segmentation problem can be rather more specific in certain application domains, like medical imaging. In such a context, regions correspond to physical objects with, in most of the cases, a geometry which is constrained by anatomical knowledge. The term knowledge-based segmentation is often employed to describe this sub-class of segmentation problems which can be implemented using two abstract concepts:

- Define the set of eligible solutions as a subspace (constrained by the anatomy, training examples, etc.) and then solve the segmentation problem in this subspace.
- Consider a model-free segmentation approach (using any of the earlier presented methods) with an additional term which penalizes the distance between the obtained solution and the expected one.

Both methods exhibit certain strengths as well as important limitations. Subspace approaches (manifold-based) have the advantage of being computationally tractable (small number of parameters to be determined) while guaranteeing that the final solution lies in the appropriate space. Active shape models [43, 46], active appearance models [42], or deformable templates [118] are some examples of such methods. These methods make certain fundamental assumptions on the manifold of the solution space and require a large training set to capture the variations of the structure under investigation. Furthermore, these methods assume an explicit description of the solution manifold. Therefore due to subspace constraints they inherit limited tolerance on deviations from the model space. Such a property can be considered as an important shortcoming in the area of medical image analysis where the main goal is to facilitate diagnosis through early detection of non-healthy subjects.

Soft (shape-based) constraints (manifold-enhanced) are an alternative to manifold-based approaches where one seeks a solution that is well supported by the data while its distance from the manifold is within an acceptable range. Parametric contours [49], deformable surfaces [172], shape-constrained level set methods [153], and more recently deformable metamorphs [82] are some examples of these methods. Their main advantage is that one can use manifolds which are not necessarily decomposable in vector sub-spaces. Therefore, one can consider various types of representations as well as different number of degrees of freedom for the modeling component towards being able to capture the observed variation of the samples. Furthermore, the data-fidelity term can be customized to the nature of the problem as well. These two terms are often integrated using some weight factor and are optimized jointly through the calculation of the partial derivatives. While these methods exhibit certain strengths they also suffer from some limitations. In theory one can consider very complex shape representations as well as sophisticated prior models which in practice are often impossible since the designed cost function is far from being convex and therefore convergence to a local minimum is highly probable. Parallel to that, one should worry about the pose estimation problem in order to be able to compare instances of the same class that have to be ”aligned”.

Therefore, a challenge exists in finding an appropriate compromise between the two approaches. Furthermore, it is important to provide not only quantitative results but also a qualitative assessment on the methods performance. In other words, the final outcome should also encode the uncertainty of the solution. In this case, one can compensate the use of reduced or more complex models by providing means of understanding the result to the end-user. To summarize, the use of reduced models leads to efficient shape representations where one can use efficient optimization strategies. At the same time, statistical models that can well express the variation of training samples are designed in accordance with the chosen representation.

The remainder of this chapter is organized as follows. First, we review model-free segmentation using as basis the active contour model presenting both the explicit and implicit representations. Then we briefly review knowledge-based segmentation using manifold-based and manifold-enhanced approaches before introducing our contributions. First, we propose a low rank model both in terms of surface modeling as well as statistical approximation of the samples. This is used to bring together our implicit thin plate spline surface modeling with the independent component analysis statistical modeling of the shapes variations introduced in the previous chapter [184]. This approach is validated using computer tomography CT scans of the left ventricle with aim to extract the myocardium surface. Then, in order to cope with more important variations as well as make the approach less organ specific we integrate the free form deformation approach [182] with a variable bandwidth statistical modeling [185]. We demonstrate the performance of the algorithm on MR

images of the corpus callosum [183]. In both cases, we use data, prior knowledge and uncertainties as explained in the previous chapter while we are able to derive local measures of confidence which integrate both shape and image support.

4.2 State of the Art in Segmentation

The segmentation problem is one of the best addressed in computer vision and medical image analysis. Existing approaches can be classified into two distinct categories: (i) model-free and (ii) model-based. Model-free methods do not make any explicit assumptions on the properties of the objects being observed in the image and aim at performing clustering through the analysis of the observed data in some feature space. Model-based methods make explicit assumptions on the object of interest either in terms of appearance or in terms of geometry or in both/joint space. Each class has its own strengths and well studied limitations. In the next sections we will briefly review model-free and model-based segmentation while providing details on approaches related to the ones presented in this chapter.

4.2.1 Model Free Approaches

Active contour model is a minimization approach to image segmentation (see 2.2.3). The shape representation is used to define different objective function components that will constrain its regularity and prevent overlapping while converging to the desired image properties. The most primitive approach using such hypotheses is the Mumford-Shah framework which partitions the image into (multiple) classes according to a minimal length curve while reconstructing the original signal in each class from the noisy observations. The problem was initially formulated in 1985 [122](solved in 1D) and generalized in 1989 [123]. The energy function is defined as:

$$E(\mathcal{C}, g) = \int_{\Omega} \|\mathcal{I}(\mathbf{x}) - g(\mathbf{x})\| d\mathbf{x} + \int_{\Omega - \mathcal{C}} \|\nabla g(\mathbf{x})\|^2 \, d\mathbf{x} + \nu |\mathcal{C}|, \tag{4.1}$$

where $\mathcal{C}$ is the boundary, $|\mathcal{C}|$ its length, and g a smooth approximation of the image $\mathcal{I}$ that may present discontinuities on $\mathcal{C}$. The discrete formulation of this function is related to the early works of Geman and Geman [70] with modified Monte Carlo methods. This was one of the first approaches to perform segmentation using the minimization of an appropriately defined cost function. The main limitation at that time was the lack of efficient mathematical and computational tools to recover the lowest potential of this function.

The 'snakes' [92], a pioneering framework introduced in 1987 by Kass, Witkin and Terzopoulos was the first efficient attempt to recover objects from images using an appropriately defined objective function. In such a context the boundary of an object was considered to be a single, smooth and connected contour. The 'snakes' consider a curve $\mathcal{C}$, parameterized by $s \in [0, 1]$. The objective function behind this model has introduced two basic assumptions on the segmentation; (i) smoothness on the solution space and (ii) boundaries between objects and the background that are determined through strong gradient values. The energy is defined as:

$$E(\mathcal{C}) = \int_0^1 \underbrace{\alpha \left\| \frac{\partial \mathcal{C}}{\partial s} \right\|^2 + \beta \left\| \frac{\partial^2 \mathcal{C}}{\partial s^2} \right\|^2}_{smoothing\ term} + \underbrace{\beta h(\mathcal{C}(s))}_{image\ term} ds$$

The smoothness term (internal) stands for regularity along the curve and has two components corresponding to the minimization of stretching and bending of the curve respectively. The image term (external) is designed in such a manner that the contour is attracted to the desired image features. The simplest form of h is a decreasing function of the image gradient norm ($h(\mathcal{C}(s)) = g(\|\nabla(\mathcal{I}(\mathcal{C}(s)))\|)$), but more advanced functionals were proposed later, using texture information, image points of interest, etc. The calculus of the variations and a local steepest gradient descend were considered to recover the lowest potential of this cost function at that time. One of the main limitations of this approach was its myopic nature, since the final solution was quite sensitive to the initial conditions. This has been addressed in the recent years using mode advanced optimization techniques including discrete approaches. However, convergence to local minimum was not the only limitation of the original method. The snake model makes an explicit assumption on the parameterization of the contour (in the most general case using control points and an interpolation function). Therefore the segmentation outcome strongly depends on the selected parameterization.

The geodesic active contour can be considered as a geometric alternative to snakes and can deal with the above mentioned limitation. It was introduced in [24, 95] and is based on a curve representation which explores the geometric properties of the contour. Let us ignore the second order smoothing term in the snake energy, then the rest of the integrals can be reformulated using the Euclidean arc length:

$$E(\mathcal{C}) = \int_0^{L(\mathcal{C})} \left\| \frac{\partial \mathcal{C}}{\partial s} \right\|^2 ds + \int_0^{L(\mathcal{C})} g(\|\nabla \mathcal{I}(\mathcal{C}(s))\|) ds,$$

where $L(\mathcal{C})$ is the length of the curve. Using Maupertuis principle and Fermat principle [24], it was proven that the minimization of the above functional is equivalent to finding the curve of minimal

length in a Riemannian space with metric being defined using the image function:

$$E(\mathcal{C}) = \int_0^L g(\|\nabla \mathcal{I}(\mathcal{C}(s))\|)ds.$$

The lowest potential of this cost function can be determined using a gradient descent method, and the computation of first order derivatives according to the Euler-Lagrange principles:

$$\frac{\partial \mathcal{C}}{\partial t} = \underbrace{g(\|\nabla \mathcal{I}(\mathcal{C})\|)\mathcal{K}\mathcal{N}}_{boundary\ force} - \underbrace{(g'(\|\nabla \mathcal{I}(\mathcal{C})\|)\mathcal{N})\mathcal{N}}_{refinement\ force}.$$

with $\mathcal{K}$ being the curvature. The first term shrinks the initial contour towards the desired image features (strong gradient) while penalizing irregularities of the curve through the curvature term. The second term is only effective close to the object boundary (where $g'(\|\nabla \mathcal{I}(\mathcal{C})\|) \neq 0$) and is used to attract the curve to the center of the boundary, which corresponds to a local maximum of the gradient norm. The use of explicit contour parameterization is computationally efficient but inherits two important limitations. One should first introduce a frequent re-parameterization of the contour to compensate local stretching and shrinking of the curve. Such an explicit approach does not allow topological changes of the initial interface. Implicit contours and level-sets approaches were introduced to overcome these constraints.

One can consider the gradient component normal to the curve (the tangential has only effect on the internal parameterization of the curve) and can express the curve evolution using the following generic formula.

$$\begin{cases} \frac{\partial \mathcal{C}}{\partial t}(s) = F(\mathcal{K}, \mathcal{C}(s), \mathcal{I})\mathcal{N} \\ \mathcal{C}(s,0) = \mathcal{C}_0(s) \end{cases}$$

where motion depends on the contour mean curvature and image data.

Implicit methods represent a closed contour (or any full shape boundary) as the 0-isolevel of an implicit function. Signed distance functions (2.2.4) are the most convenient representations both from numerical and theoretical perspectives:

$$\phi_{\mathcal{C}}(\mathbf{x}) \quad = \begin{cases} d_\Omega(x) & = ED(\mathbf{x}, \mathcal{S}) = \inf_{y \in \mathcal{S}} \|\mathbf{y} - \mathbf{x}\|, & \mathbf{x} \in \mathcal{CS} \\ -d_{\mathcal{CS}}(x) & = -ED(\mathbf{x}, \mathcal{CS}) = -\inf_{y \in \mathcal{CS}} \|\mathbf{y} - \mathbf{x}\|, & \mathbf{x} \in \mathcal{S} \end{cases}$$

The key idea of level sets is to transform the Lagrangian approach to contour evolution into Eulerian one by evolving directly the implicit representation [57, 127]. One can show that the evolution

of the contour $\mathcal{C}$ is equivalent to the evolution of the implicit function ϕ according to the scheme:

$$\frac{d\phi}{d\mathbf{x}} = -F(\mathcal{K}, \mathcal{I})|\nabla\phi|,$$

where the definition of F is extended to the entire space using the expression $\mathcal{K} = \nabla(\nabla\phi/|\nabla\phi|)$ for the curvature. This formulation handles changes of topology such as splitting and merging. However, given that the level set isophotes move with different speed functions (image-dependent) the above flow and therefore frequent updates on the distance function are to be considered to guarantee the numerical stability of the method. Level set methods have been studied both from theoretical and practical points of view [163, 129] and are among the most popular segmentation techniques. The use of level set representations had a huge impact on the further development of snake based methods.

However, the snake model itself was myopic and quite sensitive to the absence of local support. Region based segmentation was an alternative to pure edge-based snakes and developed in parallel to the level set method. In [218, 219], Zhu and Yuille have used region statistics and have proposed bayesian/minimum description length expression to estimate distributions of the different regions. In the most general case, such energy is formulated as follows:

$$E_{image}(\mathcal{I}, \mathcal{G}) \;=\; -\sum_i \int_\Omega \mathbb{1}_{\mathcal{R}_{\mathcal{C}_i}}(\mathbf{y}) \log\left(p_{\mathcal{C}_i}(\mathcal{I}(\mathbf{y}))\right) d\mathbf{y},$$

where $\{\mathcal{R}_{\mathcal{C}_1} \ldots \mathcal{R}_{\mathcal{C}_N}\}$ represent a partitioning of the image space into different regions segmenting the objects of interest. $\mathbb{1}_{\mathcal{R}_{\mathcal{C}_i}}$ is the characteristic function of the region $\mathcal{R}_{\mathcal{C}_N}$, $\mathcal{I}$ is the image and the set of $p_{\mathcal{C}_i}$ are density functions that model the statistical properties of each region. Different optimization methods and modeling of regions were proposed. The approach in [218, 219] was based on snakes and able to deal with some topological changes through a subsequent merging approach. The same objective has been achieved naturally through the use of level set methods in [25]. The Mumford-Shah functional was also used in this context for the segmentation of images into multiple regions [202]. Each region is defined with a logical combination of the different level set signs. Consequently $\log_2(n)$ levelset functions were necessary to represent n regions. The approach was extended to geodesic active contour, adding a boundary term in [130] and later extended to texture segmentation [133]. Also in [212] the approach to region based segmentation was presented to maximize the dissimilarity between the distributions defining different segmented regions. Despite improvements on the image terms, the main limitation of the deformable curves and surfaces is not related to the objective function itself but more to its optimization.

Calculus of variations systematically leads to the use of low level information (curvature, gra-

dient, intensity) and therefore the process becomes sensitive to image noise, and is likely to get stuck in a local minimum of the energy. In order to address this limitation, the first attempt was to introduce a 'balloon' force [36], which aimed to help the contour to surpass noise. On the other hand this model requires explicit knowledge of the direction of the propagation and could lead to erroneous results in the presence of weak edges. The stochastic active contours [90] is a more elegant approach to address this problem through a random perturbation of the motion, according to a simulated annealing procedure. More recently, generalized gradient flows were considered to address this limitation. These methods make an explicit use of the manifold structure of the space of curves $\mathcal{S}$, and defines the gradient locally in the tangent space of $\mathcal{S}$ using a particular metric [177]. The standard derivation method used in conjunction with gradient descent assumes that we are using the L^2 norm in the tangent space:

$$< \nabla_{\mathcal{S}} E(\mathcal{C}) | h >_{L^2} = \left. \frac{dE(\mathcal{C} + th)}{dt} \right|_{t=0},$$

where $\mathcal{C} \in \mathcal{S}$ and $h \in T_C\mathcal{S}$ defines the tangent space to $\mathcal{S}$ at position C. This approach is not restricted to explicit contours, and was directly extended to level set representations. This expression was modified with the use of a different scalar product on the tangent space. In [177] the use of Sobolev norm $W_{1,2}$ instead of L^2, was proven to be equivalent to a proper smoothing of the flow obtained with the L^2 norm. A different motivation was considered in [169, 28] leading to the same interpretations, where the authors have projected the gradient $\nabla_{\mathcal{S}} E(C)$ on different well chosen subspaces. This framework was proven to be an efficient technique to avoid local minima of the energy.

In general a method that constrains the motion of a deformable model can be interpreted as a kind of prior on the object to segment as deformation of the initial shape will be preferably rigid before going to local. Classic snake approaches can account for some global information related to the geometry of the object such as area and perimeter but cannot account for structural information. Such a task often involves recovering a probabilistic representation on the manifold of expected solutions. Once an information space describing the object of interest is available, the next task consists of penalizing image-based solutions which cannot be expressed from the prior model.

4.2.2 Model-Based Approaches

The domain of geometry-driven/aided object extraction has gained significant attention over the past decade mostly due the development of the medical imaging field. Existing methods can either be manifold based, or manifold aided. In the first case, one recovers a parametric representation of

the manifold and then imposes that the solution to be recovered in the image be part of the manifold. These methods are computationally efficient because the optimization is constrained to lie on a predefined manifold but have limited capture since any sample is to be seen before segmentation in order to be part of the model. Manifold-enhanced methods still assume a probabilistic/manifold representation of the solution space. However, these methods introduce a cost proportional to the distance between the solution and the manifold. This can be considered as an advantage since the prior model is decomposed from the segmentation solution and certain freedom exists on recovering solutions which have not been seen before. Computational complexity and computational efficiency are the main limitations of the manifold-enhanced approaches.

Manifold-Driven Object Segmentation

Deformable template fitting was the first attempt to knowledge-based object extraction in images with aim to determine the position of the eyes and mouth on faces in video sequences [214]. An image template is used to define an energy that contains image information like edges, peaks and valley of the intensity and geometric information based on the complexity of the transformation applied to the template. Optimization being performed with gradient descent, the set of parameters describing the template are used in an incremental fashion so that prominent features of the object will be detected first and then refined using all parameters. This technique is very sensitive to the initialization and was mostly tested for tracking purposes.

Active Shape Models (ASM) is one of the first known attempts to model the structural deformation of an object explicitly [45, 46]. The technique is based on a Point Distribution Model (PDM) shape representation, that is a set of points located on the feature points of the object (edges) and equally spaced to describe the contour properly. Assume the existence of a training set with manually segmented images, every shape is described with a PDM of corresponding points concatenated as a unique feature vector $\mathbf{x}$. Cootes and Taylor [46] proposed to model the variations of $\mathbf{x}$ as a weighted sum of the columns of an orthonormal matrix P using principal component analysis on the training data. Any shape of the training set can be approximated with the use of a weight vector $\mathbf{b}$:

$$\mathbf{x} = \bar{\mathbf{x}} + P.\mathbf{b},$$

where $\bar{\mathbf{x}}$ represents the mean of the samples and is used as a reference model. Towards invariance of the model with respect to the camera/sensor point of view, a registration with respect to a similarity transform was considered first, so that the above description is defined for the local deformation component. This method is using PCA to retrieve the columns of P (see 3.2.2) which only contains

the directions of the largest variation of the local deformations. Knowledge-based segmentation using this model first involves an initial estimation of the position of the object in the image, and then its local deformation using the coefficients of the modes of variations towards expressing the data. This process is done in an iterative fashion. First, given the current position of the control points, correspondences with the image (better positioning of the control points) are searched. This is done at the point level through a local search on the normal direction to the curve. The optimal correspondence is determined either using the highest gradient or through appropriate matching of intensity profiles. Once the correspondences between the model and the image for the PDM are obtained, the new pose parameters (s, R_θ, t) are computed by aligning it to the reference model, and local parameters $\mathbf{b}$ computed with the projection on the columns of P. An updated candidate shape is obtained and the process iterates until convergence, with the final result:

$$\mathbf{x}' = sR_\theta(\bar{\mathbf{x}} + P.\mathbf{b}) + \mathbf{t},$$

where by extension we have noted $R_\theta.\mathbf{x}$ the transformation of all points of the PDM with rotation R_θ. Active shape models are purely geometric ones and decorrelate geometry with appearance. Introducing appearance in the manifold was a natural extension, namely the Active Appearance Models (AAM) [42]. In this case, the training set is composed of a set of objects manually segmented (outline of the object of interest and geometric feature points of interest, etc...). Then, all extracted instances of the object are brought to the same pose according to the shape information leading to a shape parameter vector like ASM ($\mathbf{x}$). AAM is using a warping (non rigid deformation) of every shape of the training to the reference shape $\bar{\mathbf{x}}$. As shape is represented with PDM, the warping is performed with a piecewise linear transformation of the Delaunay triangulation associated with the reference shape $\bar{\mathbf{x}}$. Once all instances of the object from the training set are aligned, a feature vector is defined that contains image information (a vector $\mathbf{y}$ containing intensity of the image). The mean appearance $\bar{\mathbf{y}}$ is also computed, and the initial approach also accounted for linear contrast invariance that consists in a normalization of the histograms associated with the vectors $\mathbf{y}$. Data being pre-processed, learning possible shape and appearance variation is performed using PCA in the joint space $(\mathbf{x}, \mathbf{y})$ to reduce the dimensionality of the problem:

$$\begin{aligned}\mathbf{x}' &= sR_\theta(\bar{\mathbf{x}} + P_{ASM}.\mathbf{b}) + \mathbf{t} \\ \mathbf{y}' &= a(\bar{\mathbf{y}} + P_{AAM}.\mathbf{c}) + b,\end{aligned}$$

with $(s_{,\theta}\,, t)$ the pose parameters, (a, b) image contrast parameters and $(\mathbf{b}, \mathbf{c})$ the weights associated with the local variation of shape and appearance respectively. Segmentation is performed by deforming the model to the observed image, such that the set of parameters of the shape and image model variations reconstruct an object that looks similar to the one observed. Initial optimization

techniques were based on gradient descent iterating simultaneously on the shape and appearance parameters. Recently, the convergence properties of the approach were studied in [114] and a faster optimization method that does not require the calculus of variations was introduced in [59]. The technique was widely explored for various applications notably in segmentation of ultrasound image sequences in [17]. Active Shape and Appearance models are examples of manifold-based methods. Their main strength is computational efficiency, their main drawback shows limited applicability/capture of variations.

In order to deal with more complex models the idea of sparse models has recently been proposed [66]. It consists of recovering a basis and a linear and nonlinear interpolation strategy that satisfies three criteria, (i) good representation of the training set, (ii) optimal support from the data, (iii) invariance to local perturbations of the basis. While this contribution is showing more degrees of freedom, the retrieved segmentation is still constrained to a fixed manifold. Manifold-enhanced segmentation addresses this limitation with the use of model free approaches and penalizes the distance to a learnt shape manifold with various complexity.

Manifold-Enhanced Object Segmentation

In such a context, one would like to combine model-free and model-based approaches. This can be done through the definition of a cost function that is the sum of an image and regularisation term (inspired by model free techniques), and a shape based term, derived from a training process. In the most general case one can write:

$$E(\mathcal{C}) = E_{external}(\mathcal{C}, \mathcal{I}) + E_{internal}(\mathcal{C}) + E_{shape}(\mathcal{C})$$

These methods were first considered to improve the classic snake model. Since our approach is exploring implicit representations, we will focus on manifold-enhanced approaches in this area. The simplest possible approach relates segmentation with the Chamfer transform [31]. The prior model was rather simple: an average contour and its corresponding distance transform. In order to produce a prior term, the projection of the 0-level set (actual contour in the observed image) to the reference level-set was considered according to certain pose parameters. In order to account for the signed nature of the reference shape and to produce a nice/derivable prior the square of the projection was considered, or:

$$E(\mathcal{C}) = \int g(|\nabla \mathcal{I}(\mathcal{C}(p))|) + \frac{\lambda}{2} d^2(\mu R_\theta \mathcal{C}(p) + \mathbf{t}) \|\mathcal{C}'(p)\| dp$$

where g is a classical image term cost function, $(\mu, R_\theta, \mathbf{t})$ the pose parameters defining a similarity transform. d is the Euclidean distance between the transformed contour $\mu R_\theta \mathcal{C}(p) + \mathbf{t}$ and the reference contour $\mathcal{C}^*$. The technique requires to constantly update the pose parameters aligning $\mathcal{C}$ to the reference contour. The estimation of the pose parameters is critical in such a context since, due to the small number of observations, it can be erroneous while the prior model has also limited capture.

In [153, 154] these issues were partially addressed. In particular a more robust metric between the evolving contour and the prior was proposed while a probabilistic behavior was inherited to the reference level set model. Within this approach, a level set function $\phi^*_{\mathcal{M}}$ was considered which also involves the local scalar variance aiming to describe the possible variations of the levelset values:

$$\begin{aligned}
&\{\phi^*_{\mathcal{M}}, \sigma^*_{\mathcal{M}}\} = \arg\min_{\phi_{\mathcal{M}}, \sigma_{\mathcal{M}}} E(\phi_{\mathcal{M}}, \sigma_{\mathcal{M}}), \qquad \text{under constraint } \|\nabla \phi_{\mathcal{M}}(\mathbf{x})\| = 1 \\
&\text{with } E(\phi_{\mathcal{M}}, \sigma_{\mathcal{M}}) = \int_\Omega \sum_{i=1}^{N} -\log(p^{\mathcal{M}}_{\mathbf{x}}(\phi_i)) d\mathbf{x} \\
&\text{and } p^{\mathcal{M}}_{\mathbf{x}}(\phi) = K(\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathbf{x}), \sigma_{\mathcal{M}}(\mathbf{x})),
\end{aligned}$$

where $K(x, \sigma)$ is a Gaussian kernel with bandwidth σ valued at x. Optimization with respect to $\phi_{\mathcal{M}}$ and $\sigma_{\mathcal{M}}$ was performed in a variational framework, and these were used as a shape based energy term for the segmentation process:

$$E_{shape}(\phi) = \int_\Omega \left(log(\sigma_{\mathcal{M}}(\mathbf{x})) + \frac{(\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathbf{x}))^2}{2\sigma_{\mathcal{M}}(\mathbf{x})^2} \right).$$

This technique also requires a constant estimation of the pose parameter aligning the level set model $\phi^*_{\mathcal{M}}$ to the current segmenting level set ϕ.

Most manifold enhanced segmentation methods require to separate pose estimation from local deformation in order to design a shape prior which only accounts for local variations. Therefore pose parameters need to be estimated constantly during the registration process. This difficulty was addressed in [68], where an affine invariant shape model based on particular moments was built. Their approach implicitly computes the pose parameters that align the deformable model to a 'universal' reference frame only based on second and third order shape moments. These are used to retrieve the parameters of the affine matrix of the form:

$$A = R_\gamma D R_\theta$$

where R_γ, R_θ are rotation matrix and D a diagonal matrix. Affine invariant moments were com-

puted in a closed form directly on the initial shape. Computation of the moments up to a certain order allows reconstruction of an affine invariant reference shape. Therefore using the general theory of shape gradients [6], they have derived a shape based energy term defined as the distance between moments associated to the shape model and the deformable model respectively. They have added this term to a region based energy functional of Chan & Vese [26] to perform segmentation. These approaches have introduced robustness on imposing a prior as well as some improvement on the capture range of the model. The next step was to consider more advanced statistical models to capture the variation of a training set.

In [99] level sets and PCA manifold enhanced priors were considered in a sequential fashion. The prior model was computed using an orthogonal sub-space through a PCA on the aligned distance functions of the training examples. A mean shape model is obtained with orthogonal modes of variations:

$$\phi'(\mathbf{x}) = \bar{\phi}(R_\theta \mathbf{x} + \mathbf{t}) + \sum_{i=1}^{N} b_i \psi_i (R_\theta \mathbf{x} + \mathbf{t}),$$

where $\mathbf{b} = \{b_1 \dots b_N\}$ is the vector of local deformation parameters and $\mathbf{a} = \{R_\theta, \mathbf{t}\}$ defines the pose parameters. The levelset ϕ' is also a function of the parameters $\{\mathbf{a}, \mathbf{b}\}$. Segmentation is associated to the minimization of a function containing an image-base term based on geodesic active contour, and a shape-based term as the square distance between the current level set ϕ and the best level set candidate $\phi'(\mathbf{a}, \mathbf{b})$ approaching ϕ in the learned subspace:

$$\begin{aligned} E(\phi) &= \int \lambda_1 g(|\nabla \mathcal{I}(\mathbf{x})|) |\nabla \phi(\mathbf{x})| + \lambda_2 (\phi - \phi')^2 d\mathbf{x} \\ \phi'(\mathbf{a}', \mathbf{b}') &= \underset{\mathbf{a},\mathbf{b}}{\operatorname{argmax}} P(\mathbf{a}, \mathbf{b} | \phi, \nabla \mathcal{I}). \end{aligned}$$

The first energy term defining the evolution of the levelset $\phi(t)$ is optimized using gradient descent, while at every time step the optimal $\phi'(\mathbf{a}, \mathbf{a})$ approximating $\phi(t)$ is computed with the maximization of the second term. $P(\mathbf{a}, \mathbf{b} | \phi, \nabla \mathcal{I})$ is obtained using Bayes rule under certain independence assumptions, and leads to different elementary terms. The optimal values of $\mathbf{a}$ and $\mathbf{b}$ are retrieved by gradient ascent on the log-probability.

The work of [196] as well as the one of [155] was a natural extension of this concept leading to an implicit version of active shape or active region models. This method aimed to recover the pose parameters while evolving the contour in the image in such a way that it belongs to the orthogonal learned space while at the same time accounts for the desired image features. Despite the improvement over the static models, these methods impose a linear behavior on the training set, that is the prior manifold should be expressed using an orthogonal base. Such a linearity constraint

is not always satisfied. The use of kernel principal component analysis is an efficient tool to address this limitation.

The idea of Kernel PCA is conceptually simple: assume data living in a non-linear space, consider an infinite dimension function which projects this data to a new space where linear separation is feasible. In [47], the Mumford-Shah data term was combined with a Kernel PCA prior model. It can be proven that the data can be mapped to a high dimensional feature vector space with a non-linear function ϕ so that the kernel in data space is mapped to a dot product in feature space ($k(\mathbf{z}_i, \mathbf{z}_j) =< \phi(\mathbf{z}_i)|\phi(\mathbf{z}_i) >_F$). Using PCA in the feature space, can be formulated only with the use of a dot product [119] and therefore translates to a weighted sum of kernels in the data space. In such a case, the prior term of the energy takes the following form:

$$E(z) = \sum_k \lambda_k \sum_i w_{k,i} k(z, z_i),$$

where the $w_{k,i}$ are the components of the k^{th} eigenvector associated with eigenvalue λ_k. In [47], Cremers et al. have shown the strong relation between this approach and the Parzen window technique for non-parametric density estimation (3.2.4).

Introducing prior knowledge in segmentation involves two aspects, (i) defining an appropriate prior mode, (ii) defining a distance between the current solution and the model. Most of prior art involves Euclidean distances between shapes which can be considered as a limitation. Recently, more complex distances were considered like Sobolev and Hausdorf [27] and the manifold structure of the shapes was considered. Using linear methods like PCA and ICA in [28], the prior was built under the assumption of small shape deformations, and defined in the tangent space associated with the mean shape.

Meanwhile, manifold-based dimensionality reduction techniques were adapted to the set of shapes and used for segmentation. Such methods assume that the set of shapes of the training is sampled from a smooth finite dimensional sub-manifold of the manifold of shapes. These techniques are twofold, first estimating a low dimension smooth manifold from the sampled data and secon building an embedding from this manifold to a finite dimensional Euclidean space. For instance, in [64], Laplacian eigenmaps and diffusion maps were considered. This embedding is used as a basis to build a continuous manifold out of the limited set of training samples. The second step consists of computing the projection of any shape on this manifold. The shape prior term can be designed as the minimization of the distance from any shape to the learned shape manifold.

Let us now consider an axiomatic approach to knowledge-based segmentation. The properties

of such an approach or constraints can be summarized as follows: (i) a model that can account for the variations of the training examples and can go beyond that, (ii) a model that is as compact as possible and involves a small number of parameters, (iii) a constraint that explores the model towards recovering solutions that are consistent with the model, (iv) a method that uses the image features as optimally as possible. The answer to the last condition is rather trivial since in the most general case region-based techniques and separation of classes using statistics on intensities is the most appropriate approach. This is not the case for the first three conditions. Inspired from our work on modeling shape variations with uncertainties, the selection of the most appropriate model is natural. Problems which involve samples of limited variation can be captured with linear models which encode uncertainties while non-linear structures like the human brain can be modeled using non-parametric models and kernels. In terms of nature of constraint, since dimensionality reduction has been already addressed in the model space, we consider a manifold-enhanced approach where the optimization is addressed through deformation of the reference shape according to the image and the prior on the deformation space.

4.3 Manifold-Enhanced Knowledge-Based Segmentation

A straightforward approach to the problem of image segmentation is based on the Bayesian formulation of the maximum a posteriori. This model allows the natural use of intensity, texture region information along with a shape prior [132, 128]. Let $\mathcal{P}(\Omega)$ represent a partition of an input image $\mathcal{I}$. The density function of the partitioning given the image which we want to maximize according to the Bayes rule is given by:

$$p(\mathcal{P}(\Omega)|\mathcal{I}) = \frac{p(\mathcal{I}|\mathcal{P}(\Omega))p(\mathcal{P}(\Omega))}{p(\mathcal{I})},$$

where:

- $p(\mathcal{I}|\mathcal{P}(\Omega))$ is the posterior probability of the image $\mathcal{I}$ given the segmentation result. This term measures how well the segmentation $\mathcal{P}(\Omega)$ explains the image, in terms of homogeneity of intensity, distribution or texture information. The choice of this term is critical and should motivate the feature space where separation between classes is straightforward.
- $p(\mathcal{P}(\Omega))$ is the prior probability of the segmentation. This term may include smoothness constraints on the segmented regions, as well as the statistical shape prior.
- $p(\mathcal{I})$ is the probability of having the image $\mathcal{I}$ in the space of all images. This term which do

not contain any information is usually considered as a constant normalization term.

Recovering the sample that corresponds to the highest peak of the above probability density is equivalent to the minimization of $-\log(p(\mathcal{P}(\Omega)|\mathcal{I}))$, this leads to the classical energy based approaches to image segmentation:

$$E(\mathcal{P}(\Omega)) = -\log(p(\mathcal{P}(\Omega)|\mathcal{I})) = \underbrace{[-\log(p(\mathcal{I}|\mathcal{P}(\Omega)))]}_{Image\ term} + \underbrace{[-\log(p(\mathcal{P}(\Omega)))]}_{Internal\ +\ shape\ prior\ terms}$$

The definition of these components should be application driven and will be discussed in the upcoming sections. However, even if the descriptors are aiming to separate the object from the background, their definition may vary from one application to another, the theoretical model (region-based segmentation) is the same.

4.3.1 Image-Based Term

Without loss of generality let us assume that the prior model consists of a Full Shape or a set of non-overlapping full shapes $\mathcal{S}_i$ as defined in section (2.2). In the latter case, this defines a multi-components shape model. Then, we can go even further and assume that this shape model is associated with a template related to the different structures being present in the image. Section 2.3 has defined a set of deformations applied to the shapes and having the entire domain Ω as support. Deforming the shapes is therefore equivalent to deforming the associated template.

Let us now assume n-classes being part of the model ($\{\mathcal{C}_1 \dots \mathcal{C}_n\}$) + a background class ($\mathcal{C}_0$). Let $\mathbf{x}$ be a point of the template, then $\zeta(\mathbf{x})$ is the class of the point $\mathbf{x}$.while In the most general case, the template is probabilistic: the probability of any point of the template belonging to class $\mathcal{C}_i$ is $p(\zeta(\mathbf{x}) = \mathcal{C}_i)$, therefore satisfying $\sum_{i=0}^{n} p(\zeta(\mathbf{x}) = \mathcal{C}_i) = 1$. In case the template is determinisitic, then $p(\zeta(\mathbf{x}) = \mathcal{C}_i) \in \{0, 1\}$, and we can define regions on the template associated with each class ($\mathcal{R}_{\mathcal{C}_i} = \{\mathbf{x}|Class(\mathbf{x}) = \mathcal{C}_i\}$), which also correspond to the reference shapes $\mathcal{S}_i$.

Consequently, one can define the model free segmentation problem through a warping of the reference shapes (and the template being associated with it) in a very general fashion with the maximum a-posteriori. One will retrieve the principal mode of $p(\mathcal{I}|\mathbf{\Theta}, \mathcal{G})$, where $\mathbf{\Theta}$ is the vector defining parametric transformation of the template and $\mathcal{G}$ refers to the visual properties of each

class:

$$\boldsymbol{\Theta}_{max} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \left(p(\mathcal{I}|\boldsymbol{\Theta}, \mathcal{G}) \right)$$
$$p(\mathcal{I}|\boldsymbol{\Theta}, \mathcal{G}) = \prod_{\mathbf{y} \in \Omega_{\mathcal{I}}} \left(\sum_i p(\mathcal{I}(\mathbf{y})|\zeta(\mathcal{L}^{-1}(\boldsymbol{\Theta}, \mathbf{y})) = \mathcal{C}_i, \boldsymbol{\Theta}, \mathcal{G}).p(\zeta(\mathcal{L}^{-1}(\boldsymbol{\Theta}, \mathbf{y})) = \mathcal{C}_i|\boldsymbol{\Theta}) \right)$$

where $\mathcal{L}$ is the parametric transform associated with the vector of parameters $\boldsymbol{\Theta}$.

In practice the class probability of any point of the template can be determined through a learning process. In our case, the reference shapes defining the object of interest have been aligned to a training set of pre-segmented corresponding shapes with respect to the parametric transformation $\mathcal{L}$. Thanks to this shape registration process, the separation between classes is very well defined and the associated template can be considered as nearly deterministic. Therefore, the class probability $p(Class(\mathbf{x}) = \mathcal{C}_i)$ will be approximated with the use of smoothed Heaviside function showing fast transition of the class probability of $\mathcal{C}_i$ from 1 to 0 when crossing the boundary defined by the shape $\mathcal{S}_i$.

$$p(Class(\mathbf{x}) = \mathcal{C}_i|\mathcal{G}) = p(Class(\mathbf{x}) = \mathcal{C}_i) = H_\sigma \left(\phi_{\mathcal{R}_{\mathcal{C}_i}}(\mathbf{x}) \right),$$
$$with \qquad H_\alpha(t) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left(\frac{t}{\alpha} \right)$$

where $\phi_{\mathcal{R}_{\mathcal{C}_i}}$ is the signed distance computed from the class $\mathcal{C}_i$ on the template, and σ is the smoothing parameter of the Heaviside function. One has also noticed that the class probability is naturally independent of the visual property of each class.

Now consider that the transformation vector is known, one can write the probability of an image point $\mathbf{y}$ belonging to the class $\mathcal{C}_i$, conditional to the transformation vector $\boldsymbol{\Theta}$:

$$p(\zeta(\mathbf{y}) = \mathcal{C}_i|\boldsymbol{\Theta}) = H_\alpha \left(\phi_{\mathcal{R}_{\mathcal{C}_i}}(\mathcal{L}^{-1}(\boldsymbol{\Theta}, \mathbf{y})) \right),$$

In order to facilitate the notation, let p_{C_i} be the distribution of the greylevel intensity for the class C_i. Then, segmentation is equivalent of maximizing the a-posteriori:

$$\begin{aligned} p(\mathcal{I}|\boldsymbol{\Theta}, \mathcal{G}) &= \prod_{\mathbf{y} \in \Omega_{\mathcal{I}}} \left(\sum_i p_{\mathcal{C}_i}(\mathcal{I}(\mathbf{y})).p(Class(\mathbf{y}) = \mathcal{C}_i|\boldsymbol{\Theta}) \right) \\ &= \prod_{\mathbf{y} \in \Omega_{\mathcal{I}}} \left(\sum_i p_{\mathcal{C}_i}(\mathcal{I}(\mathbf{y})).H_\alpha \left(\phi_{\mathcal{R}_{\mathcal{C}_i}}(\mathcal{L}^{-1}(\boldsymbol{\Theta}, \mathbf{y})) \right) \right). \end{aligned}$$

where $\mathcal{L}$ refers to the parametric transform associated with the vector of parameters $\mathbf{\Theta}$ and $\mathcal{L}^{-1}$ its inverse. This expression defines the likelihood of the image given the model, the product considers all the voxels in the image. It is useful to define and minimize this quantity with the use of the log-likelihood instead, so that the segmentation can be formulated as an energy minimization problem. The particular form of the Heaviside function allows to write the approximation of the actual log-likelihood by letting the factors H_α out of the logarithm:

$$E_{image}(\mathcal{I}, \mathbf{\Theta}, \mathcal{G}) = -\int_{\Omega_\mathcal{I}} \sum_i \log\left(p_{\mathcal{C}_i}(\mathcal{I}(\mathbf{y}))\right) . H_\sigma\left(\phi_{\mathcal{R}_{\mathcal{C}_i}}(\mathcal{L}^{-1}(\mathbf{\Theta}, \mathbf{y}))\right) d\mathbf{y} \tag{4.2}$$

In the limit case when $\alpha \to 0$, this energy is simply expressed as the sum of the likelihood of the different disjoint regions:

$$E_{image}(\mathcal{I}, \mathbf{\Theta}, \mathcal{G}) = -\sum_i \int_{\mathcal{L}(\mathbf{\Theta}, \mathcal{R}_{\mathcal{C}_i})} \log\left(p_{\mathcal{C}_i}(\mathcal{I}(\mathbf{y}))\right) d\mathbf{y}$$

The optimal value for $\mathbf{\Theta}$ is obtained through gradient descent. We have computed the derivative of the expression with respect to the parameter vector:

$$\frac{dE_{image}(\mathcal{I}, \mathbf{\Theta}, \mathcal{G})}{d\mathbf{\Theta}} = -\int_{\Omega_\mathcal{I}} \sum_i \log\left(p_{\mathcal{C}_i}(\mathcal{I}(\mathbf{y}))\right) . \frac{dH_\alpha\left(\phi_{\mathcal{R}_{\mathcal{C}_i}}(\mathcal{L}^{-1}(\mathbf{\Theta}, \mathbf{y}))\right)}{d\mathbf{\Theta}} d\mathbf{y}$$

Details about this energy derivation can be found in Appendix A. It is performed using continuous expression of the Heaviside function, where one considers the limit case $\alpha \to 0$ corresponding to an exact partitioning of the template. The obtained expression uniquely considers the boundaries separating different classes of the template:

$$\frac{dE_{image}(\mathcal{I}, \mathbf{\Theta}, \mathcal{G})}{d\mathbf{\Theta}} = -\sum_{i \neq j} \int_{\partial\Omega_{\mathcal{C}_i} \cap \partial\Omega_{\mathcal{C}_j}} L_{\mathcal{C}_i/\mathcal{C}_j}(\mathbf{x}) . \nabla\phi_{\mathcal{L}(\mathbf{\Theta}, \mathcal{R}_{\mathcal{C}_i})}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x})) \frac{\partial\mathcal{L}}{\partial\mathbf{\Theta}}(\mathbf{\Theta}, \mathbf{x}) d\mathbf{x}, \tag{4.3}$$

where $\phi_{\mathcal{L}(\mathbf{\Theta}, \mathcal{R}_{\mathcal{C}_i})}$ is the distance transform of the region $\Omega_{\mathcal{C}_i}$ transformed with the parametric transformation $\mathcal{L}(\mathbf{\Theta}, .)$. The gradient of this distance is evaluated at the transform points $\mathbf{x}' = \mathcal{L}(\mathbf{\Theta}, \mathbf{x})$ of the regions boundary and weighted with $L_{\mathcal{C}_i/\mathcal{C}_i}(\mathbf{x})$ to act as a local force of the boundary. Consequently, the term $\partial\mathcal{L}(\mathbf{\Theta}, \mathbf{x})/\partial\mathbf{\Theta}$ projects the local influence on the vector of parameters. Last, $L_{\mathcal{C}_i/\mathcal{C}_j}(\mathbf{x})$ is the difference of the log-likelihood of two regions sharing an interface:

$$L_{\mathcal{C}_i/\mathcal{C}_j}(\mathbf{x}) = \log\left[p_{\mathcal{C}_i}(\mathcal{I}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x})))\right] - \log\left[p_{\mathcal{C}_j}(\mathcal{I}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x})))\right]$$

Once the image term has been introduced, the next step consists of imposing constraints on the

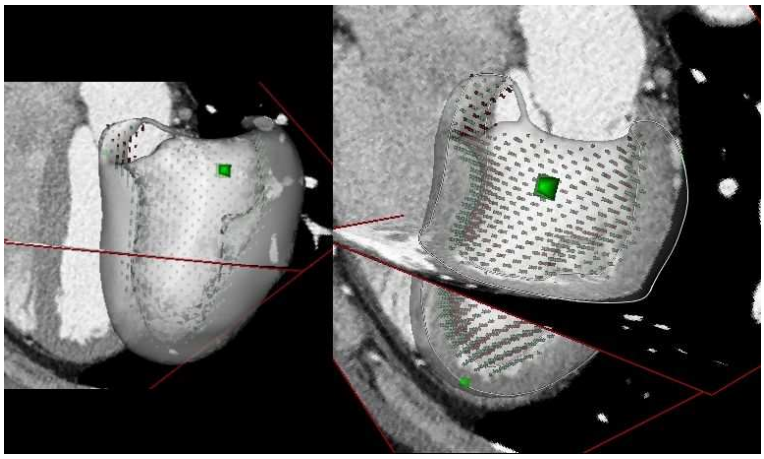


Fig. 4.1: Image Based segmentation of the cardiac left ventricle in CT-images, constraining the transformation to a similarity transform and considering uniquely the endocardium (interface between blood pool and muscle).

deformation model. In the context of our approach, segmentation is done through the deformation of the template being associated with the reference surface. In other words, the unknown parameters of the approach is the transformation of the template. In chapter 3, we have discussed methods to recover a compact probabilistic model for these transformations using the training set.

4.3.2 Shape-Based Term

Let us consider an estimator $\hat{f}(\mathbf{\Theta})$ for the probability density of deformation parameters $\mathbf{\Theta}$ describing the object of interest. Using the Bayesian formulation of the posterior density (4.3), the optimal solution is the one minimizing the sum of an image-based term and a shape prior term aiming to impose consistency with the shapes observed in the training set. Such energy is defined as:

$$E(\mathcal{A}, \mathbf{\Theta}) = \alpha E_{image}(\mathcal{A}, \mathbf{\Theta}) + (1-\alpha) E_{shape}(\mathbf{\Theta}) = \alpha E_{image}(\mathcal{A}, \mathbf{\Theta}) - (1-\alpha)\log(\hat{f}(\mathbf{\Theta}))$$

with α being a blending parameter or a normalization factor between the two term, '$\mathcal{A}$' accounting for global transformation and '$\mathbf{\Theta}$' for local transformation. Let us now consider the two statistical models of variation presented in chapter 3.

In the first case, the shape based energy term makes use of the prior density function based on Independent Component Analysis. A prior probability density function was defined on the entire space with the use of 1D Gaussian mixture fitting for every independent component (see section

3.6). We should recall that only the local deformation of the transformation was modeled, and therefore a need exists to make the method invariant to similarity. The current similarity carried by the TPS is extracted from the general form using the same approach as presented in subsection (2.3.2). The inverse of this transformation is subsequently applied to the control points to remove the pose parameters associated with the current segmentation and leads to the similarity invariant parameter vector $\mathbf{\Theta}'$.

The shape prior energy term is equivalent to the maximization of the log likelihood of the current vector of parameter $\mathbf{\Theta}$.

$$E_{\text{shape}}(\mathbf{\Theta}) = -\sum_{i=1}^{m} \log(p_i([\mathbf{W}.\mathbf{\Theta}']_i)), \tag{4.4}$$

where $\mathbf{W}.\mathbf{\Theta}'$ refers to the projection of the current shape on the independent components, p_i refers to the density estimated in (eq. 3.6) and (eq. 3.7) for the i^{th} ICA component.

Let us now consider the non-parametric prior case as developed in section (3.5). Similar to the case of the ICA, the optimal segmentation solution corresponds to the lowest potential of:

$$E(\mathcal{A}, \mathbf{\Theta}) = \alpha E_{image}(\mathcal{A}, \mathbf{\Theta}) - (1-\alpha)\log(\hat{f}_S(\mathbf{\Theta}))$$

In this expression the shape term is based on the *sample point* kernel estimator $\hat{f}_S$ and only accounts for local deformations while the image term composes global and local transformation.

Our segmentation framework allows to compute uncertainty during the evolution of the deformed shape model and consequently use *hybrid* kernel estimator. Uncertainty is expressed under the form of a covariance matrix and computed following the approach introduced in section (2.5). In this case uncertainty may only account for the linear structure of the transformed model and therefore allows variations of $\mathbf{\Theta}$ that create tangential displacements of the contour. It is computed as in equation (2.52) where the distance map of the target shape is replaced with the distance map of the transformed shape:

$$\Sigma_{\mathbf{\Theta}}^{-1} = \frac{1}{\beta}\int_{\partial\mathcal{S}} \mathcal{X}(\mathbf{x})^T \nabla\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R})}(\mathbf{x}')\nabla\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R})}(\mathbf{x}')^T \mathcal{X}(\mathbf{x}) d\mathbf{x}$$

where $\mathbf{x}' = \mathcal{L}(\mathbf{\Theta}, \mathbf{x})$ and $\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R})}$ is the signed distance map computed from the transformed region $\mathcal{L}(\mathbf{\Theta}, \mathcal{R})$. One notices that this expression only accounts for local deformations as global transformation has been subtracted from the statistical shape model. The gradient $\nabla\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R})}$ admits a closed form solution based on the Jacobian of the transformation and the gradient to the initial

shape model (see Appendix A):

$$\nabla\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R})}(\mathbf{x}') \propto \mathrm{com}\left[\frac{d}{d\mathbf{x}^T}(\mathcal{L}(\mathbf{\Theta},\mathbf{x}))\right]^T.\nabla\phi_{\mathcal{R}}(\mathbf{x})$$

where 'com' denotes the matrix of cofactors. Uncertainty being scaled according to section (2.5.5), it can now be used in the gradient descent with *hybrid* kernel estimator as defined in equation (3.20):

$$E_{shape}(\mathbf{\Theta},\Sigma_{\mathbf{\Theta}}) \;=\; -\log(\hat{f}_G(\mathbf{\Theta},\Sigma_{\mathbf{\Theta}})) \tag{4.5}$$

The use of *hybrid* estimators smooths out the sample point estimator according to local uncertainty estimate.

4.3.3 Segmentation with Uncertainties

Medical imaging is an area where errors are critical. Furthermore, computer-aided diagnosis should not aim at replacing humans but should be a tool that helps the physicians to make more appropriate decisions. Therefore methods being able to determine the quality of the obtained solution are required. As stated earlier, this can be done using the uncertainties of the process that relate the Hessian matrix of the objective function to the variance of the distribution describing the space of solutions.

To validate the segmentation results in a quantitative manner we propose to estimate uncertainties on the deformation, expressed on the position of TPS control points. These uncertainties are actually based on the structure matrix of the image-term of the energy and the normal to the segmenting contour. Such local information are subsequently projected on the vectors of parameters $\mathbf{\Theta}$. In practice, uncertainties are expressed in the same fashion as (eq. 2.52) where the local shape information part $[\nabla\phi_{\mathcal{T}}(\mathbf{x}').\nabla\phi_{\mathcal{T}}(\mathbf{x}')^T]$ is replaced with a symmetric matrix which accounts for the visual separation of the classes:

$$\Sigma_{\mathbf{\Theta}}^{-1} = \frac{1}{\beta}\int_{\partial\mathcal{S}} \mathcal{X}(\mathbf{x})^T\Sigma^+(\mathbf{x})^T\mathcal{X}(\mathbf{x})d\mathbf{x}$$

$$\Sigma^+ = \frac{\|\mathbf{a}\|}{\|\mathbf{b}\|}.\mathbf{b}.\mathbf{b}^T + \frac{\|\mathbf{b}\|}{\|\mathbf{a}\|}.\mathbf{a}.\mathbf{a}^T \qquad \text{with} \begin{cases} \mathbf{a} = \nabla(L_{\mathcal{C}_i/\mathcal{C}_j}(\mathbf{x}')) \\ \mathbf{b} = \nabla(\phi_{\mathcal{S}'}(\mathbf{x}')) \\ L_{\mathcal{C}_i/\mathcal{C}_j}(\mathbf{x}') = \log\left[p_{\mathcal{C}_i}(\mathcal{I}(\mathbf{x}'))\right] - \log\left[p_{\mathcal{C}_j}(\mathcal{I}(\mathbf{x}'))\right] \\ \mathbf{x}' = \mathcal{L}(\mathbf{\Theta},\mathbf{x}) \end{cases} \tag{4.6}$$

where integration is performed along the interfaces considered for segmentation. The global behavior of such uncertainties is described as follows: The contributions to the uncertainty along the shape surface are $\nabla\phi_{\mathcal{S}'}(\mathbf{x}')$ the normal to the transformed shape and $\nabla(L_{\Omega_\mathrm{i}/\Omega_\mathrm{j}}(\mathbf{x}'))$ which represents the quality of region separation on the interface (accounts for the visual support). The principal eigenvectors of Σ are the first bisector of $\mathbf{a} = \nabla(L_{\Omega_\mathrm{i}/\Omega_\mathrm{j}}(\mathbf{x}')))$ and $\mathbf{b} = \nabla\phi_{\mathcal{S}'}(\mathbf{x}'$ associated with the eigen value $\|\mathbf{a}\|\|\mathbf{b}\|(1+\cos(\hat{\mathbf{a}},\mathbf{b}))$. The other eigenvector is the second bisector of $\mathbf{a}$ and $\mathbf{b}$ associated to the eigenvalue $\|\mathbf{a}\|\|\mathbf{b}\|(1-\cos(\hat{\mathbf{a}},\mathbf{b}))$. Consequently, if the direction of these two vectors match exactly, then the uncertainty is concentrated in the tangential directions to the shape boundary. If these directions differ, uncertainties are also present in the normal direction to the shape.

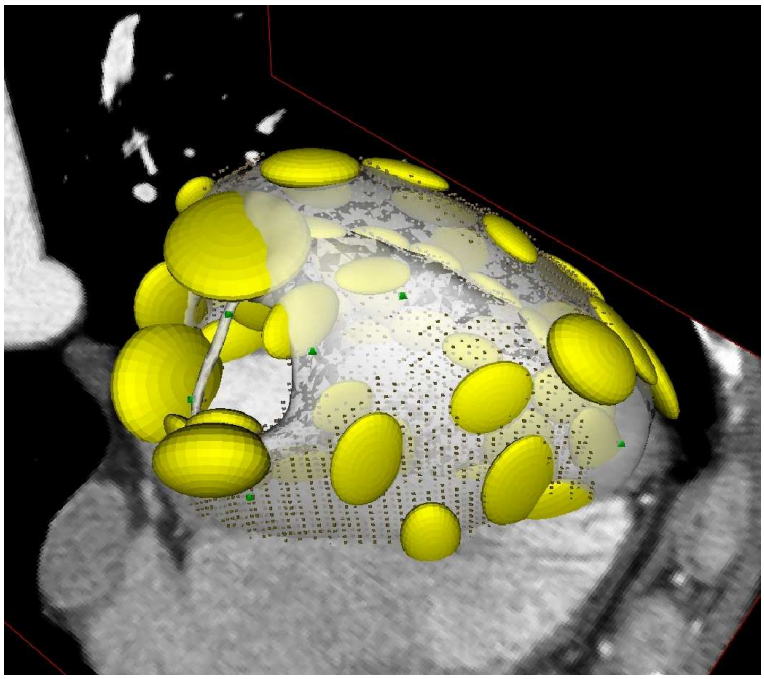


Fig. 4.2: Uncertainty computed on the segmentation result of cardiac left ventricle. Representation of its projection on the control points as a set of ellipsoids.

This local contribution is integrated along the segmented interface and projected on the parameter space (eq. 2.52) to define the segmentation uncertainty information as presented in figure (4.2). However the contribution of prior knowledge in the uncertainty and its use as guidance for the segmentation process is still to be explored. We will now describe the two medical applications of our knowledge based segmentation approach. Focusing first on the segmentation of the left ventricle.

4.3.4 Left Ventricle Segmentation from CT-scans

The a-posteriori distribution of the greylevel given the class of a voxel takes advantage of the Hounsfield units in the case of the 3D CT-scan of the heart. The classes corresponding to the Blood Pool (BP) and Left Ventricle muscle (MC) have a single monochromatic distribution. The background component shows a more complex mixture due to the presence of different tissues and organs (lung, diaphragm, atria, right ventricle) (see Fig. 4.3).

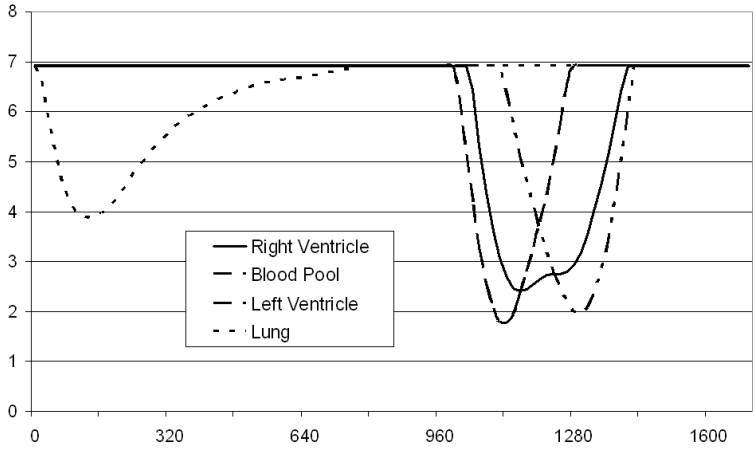


Fig. 4.3: The histograms of the greylevels of different areas of the myocardium.

In the particular case of the background class, we have split this area on the template into different sub-regions with simpler greylevel distributions, therefore separating at the template level the positions of the lung (L), Right Ventricle (RV), and everything remaining (RM). We have estimated the distributions of these different regions and applied it to our model. However the initial template was built only considering the interfaces $(BP) \leftrightarrow (MC)$ and $(MC) \leftrightarrow (Bkg)$. Therefore we do not have any prior knowledge on the position of the lung and right ventricle when deforming the template. In order to assess the likelihood of an image voxel located in the background class, a natural way would consider the maximum value of the intensity response for the different sub-classes:

$$\log\left[p_{Bkg}(\mathcal{I}(\mathcal{L}(\mathbf{\Theta},\mathbf{x})))\right] = \max_{\mathcal{C}_i=\{L,RV,RM\}} \log\left[p_{\mathcal{C}_i}(\mathcal{I}(\mathcal{L}(\mathbf{\Theta},\mathbf{x})))\right]$$

This approach is not sufficient due to the similarity between the distribution of the left ventricle muscle (MC) and right ventricle (RV), voxels classified with this maximum a posteriori probability are most often misclassified, and such method fail to converge. Therefore we have used a weak prior knowledge on the positions of the lung and right ventricle. Considering the template model,

we have isolated on the epicardium the surfaces are corresponding to the interfaces with the lung and the right ventricle. The positions of the two organs are based on the initial volume that was used to create the shape reference template (see section 3.3.1) onto which lung and right ventricle were segmented. We have reduced the size of the interfaces (left ventricle $\leftrightarrow$ lung) and (left ventricle $\leftrightarrow$ right ventricle) on the template so that these two interfaces match the actual interface (LV $\leftrightarrow$ L) and (LV $\leftrightarrow$ RV) of the image to be segmented after deformations of the template.

Therefore we have transformed the derivation of the energy term with the use of the lung and right ventricle intensity distributions:

$$\frac{dE_{image}(\mathcal{I},\mathbf{\Theta},\mathcal{G})}{d\mathbf{\Theta}} = -\sum_{(\mathcal{C}_i,\mathcal{C}_j)\in\mathcal{H}} \int_{\partial\mathcal{R}_{\mathcal{C}_i}\cap\partial\mathcal{R}_{\mathcal{C}_j}} L_{\mathcal{C}_i/\mathcal{C}_j}(\mathbf{x}).\nabla\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R}_{\mathcal{C}_i})}(\mathcal{L}(\mathbf{\Theta},\mathbf{x}))\frac{\partial\mathcal{L}}{\partial\mathbf{\Theta}}(\mathbf{\Theta},\mathbf{x})d\mathbf{x},$$

$$\text{with} \qquad \mathcal{H} = \{(\mathrm{LV},\mathrm{L});(\mathrm{LV},\mathrm{RV});(\mathrm{LV},\mathrm{BP})\}$$

We use standard estimates from the training set for the densities $p_{\mathrm{BP}}, p_{\mathrm{MC}}, p_{\mathrm{L}}, p_{\mathrm{RV}}$. Initially, a rough segmentation is performed by aligning the model with respect to a similarity transform where uniquely the interface between left ventricle myocardium and blood pool regions are considered (see figure 4.1), taking advantage of the high contrast between blood and muscle. Then we consider local deformations with the use of Thin Plate Spline with 90 control points (see section 2.3.2). The 4 classes of the complete model are used to adjust the position of the epicardium. On top of this deformation model based on image intensity, prior knowledge is used following the ICA model introduced earlier. Subsequently, segmentation is performed in 3 successive steps using gradient descent on the global energy $E = E_{image} + \alpha E_{\mathrm{shape}}$. First, image based ($\alpha = 0$) affine segmentation of the endocardium, then image based segmentation of the entire model of left ventricle with 60 control points, then segmentation with prior of the entire ventricle with 90 control points. Fig. 4.4 shows that we obtain very promising results with this method. For the first 3 examples, the average error is around 1.5 mm and the maximum error is less than 5mm. The errors are larger for the last example because this patient has a thick layer of fat around the heart and the method mistakenly combines the fat with the myocardium. In the future we are planning to broaden the lung class with a fat class.

4.3.5 Segmentation of the Corpus Callosum

In a similar fashion, we have applied our method to the segmentation of the corpus callosum in MR mid-sagittal brain slices. A bimodal partition of the image domain is to be recovered, separating the corpus callosum (cor) essentially composed of white matter from the background (bkg). A single

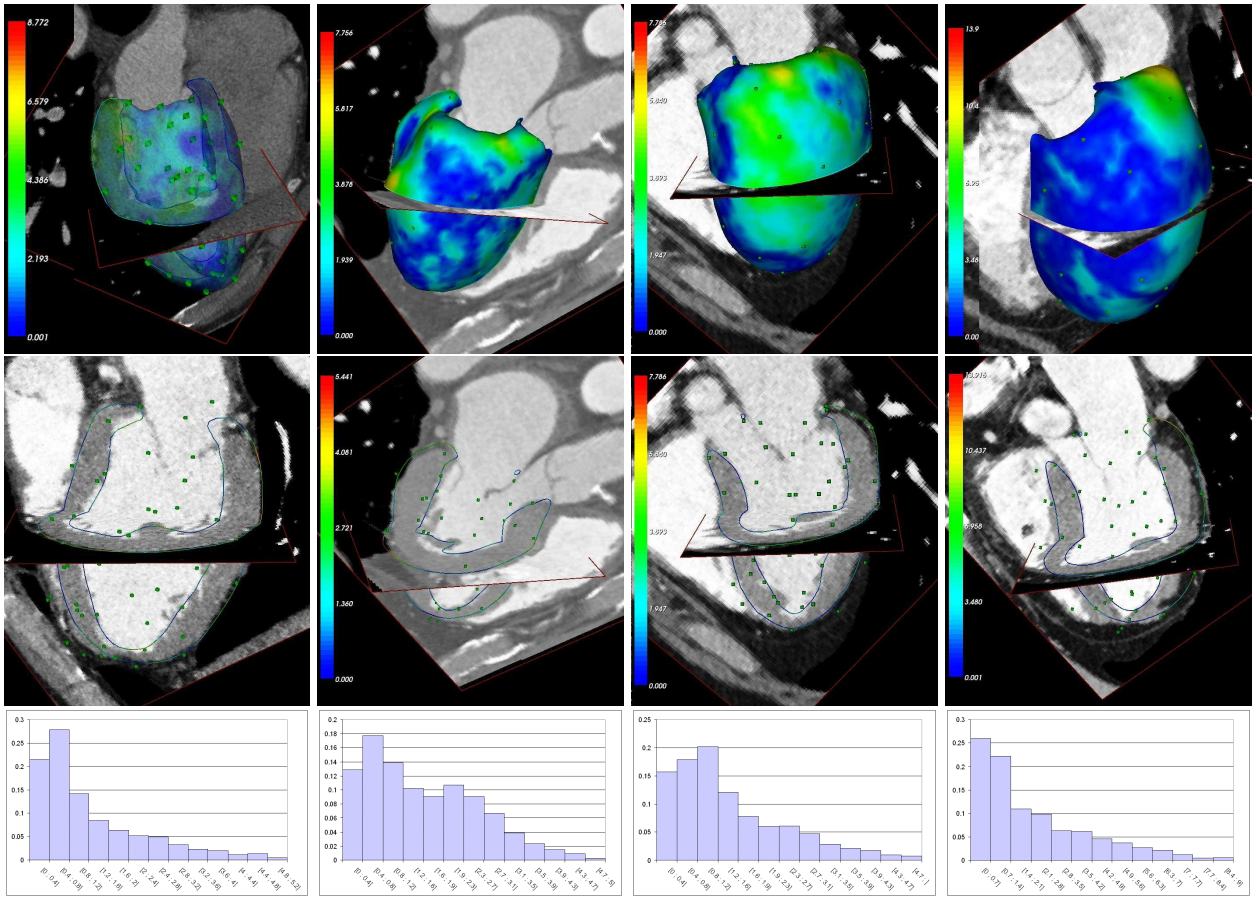


Fig. 4.4: Results of the segmentation process. Left: Segmentation of the myocardium displaying the distance to the groundtruth along with a colormap considering four different sample cases outside of the learning database. All results are displayed for a TPS deformation using the ICA shape model with 90 control points. Middle column: The intersection of the segmentation with the data, papillary muscles are correctly segmented. Right: Distribution of errors displayed as histograms. All numerical values are expressed in millimeters.

region is sufficient to represent the model in the space of the template. Consider $\mathcal{R}_{cor}$ this region, the segmentation problem consists of globally and locally deforming $\mathcal{R}_{cor}$ towards delineating the corpus callosum in the image $\mathcal{I}$. The deformation model makes use of an affine transformation and Free Form Deformations (see section 2.3.2). Therefore the problem needs to be formulated with the composition of an affine transform $\mathcal{A}$ and a local FFD $\mathcal{L}(\mathbf{\Theta}, .)$:

$$E_{image}(\mathcal{A}, \mathbf{\Theta}) = -\int_{\mathcal{R}_{cor}} \log\left[p_{cor}\left(\mathcal{I}\left(\mathcal{A} \circ \mathcal{L}(\mathbf{\Theta}; \mathbf{x})\right)\right)\right] d\mathbf{x} - \int_{\Omega - \mathcal{R}_{cor}} \log\left[p_{bkg}\left(\mathcal{I}\left(\mathcal{A} \circ \mathcal{L}(\mathbf{\Theta}; \mathbf{x})\right)\right)\right] d\mathbf{x} \tag{4.7}$$

In practice the distributions of the corpus callosum as well as the ones of the surrounding region $[p_{cor}, p_{bkg}]$ can be recovered in an incremental fashion using the Mumford-Shah principle [124]. To do so, each distribution is estimated by fitting a mixture of Gaussians to the image histograms

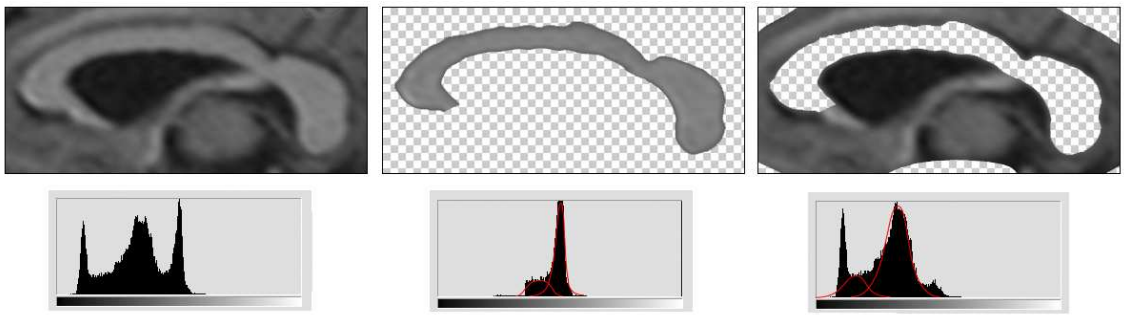
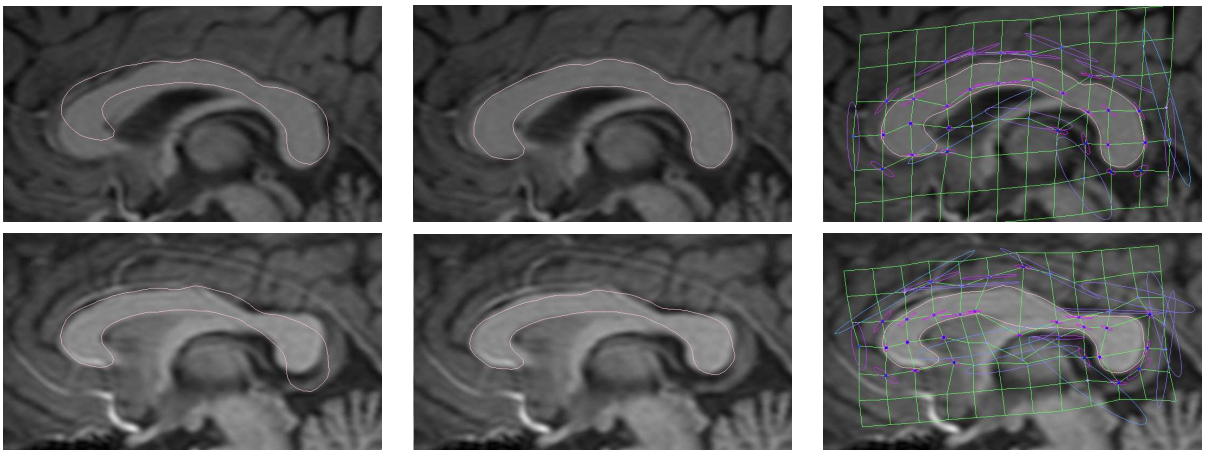


Fig. 4.5: Histograms of the corpus callosum and the background area. The of a use gaussian mixture to model the corpus callosum and background intensity distribution in MR is appropriate (this figure should be seen in color).

using the Expectation-Maximization algorithm. Fig. 4.5 shows the histogram of a typical image of the corpus callosum. The figure illustrates how well Gaussian mixtures can represent the individual histograms for the corpus callosum and the background, respectively. Histograms of the segmented image are built considering the transformed template. We use as intensity samples of the corpus callosum the inside of the transformed region $\mathcal{R}_{cor}$ and estimate the distribution with a single Gaussian distribution. We use as intensity samples of the background the transformation of a large band surrounding the outside of $\mathcal{R}_{cor}$, the obtained distribution is approximated with a mixture of 2 Gaussians.

The segmentation process is initialized by positioning the initial contour according to the method proposed in [103]. Energy minimization is performed using standard gradient descent where derivation is computed with respect to the affine transformation and the vector of parameters $\mathbf{\Theta}$. Minimization is also alternating between descent in the transformation parameters, and estimation of regional distributions. An initial gradient descent is performed, where optimization with respect to an affine transformation is considered to retrieve a rough segmentation of the corpus callosum. In a second step, free form deformation on a $[7 \times 12]$ lattice is used to retrieve the outline of the object. At this point image contrast information is not sufficient to capture the actual outline of the corpus callosum, Therefore a shape prior is essential. Following the kernel based approach earlier introduced, uncertainty in the transformed shape is computed at every step of the optimization and used to build the hybrid kernel estimator of the shape-based energy term. Segmentation results are presented in figure (4.6) along with the associated uncertainties. We demonstrate the individual steps of the segmentation process: the left most image shows the automatic initialization of the contour, the middle image shows the contour after the affine transformation has been recovered, and the right image shows the local deformations.



(a) (b) (c)

Fig. 4.6: Segmentation with uncertainties estimates of the corpus callosum; (a) Automatic rough positioning of the model, (b) segmentation through affine transformation of the model (c) segmentation using the local deformation of the FFD grid and uncertainties estimates on the registration/segmentation process (this figure should be seen in color).

4.4 Conclusion

This chapter has developed the application to segmentation of our statistical shape models. Segmentation of the corpus callosum and the segmentation of the left ventricle in cardiac CT images have been used to demonstrate the potentials of our approach. Within these approaches based on deformable models, we have introduced computation of uncertainty on the deformed model and used it into the statistical shape model. Also we have proposed a model for the computation of uncertainty in the visual features of the segmented image.

Future work may investigate a better understanding of uncertainty computation, generalizing the framework proposed for shape registration. This axis of research is particularly important as applications to medical assisted diagnosis are immediate. Finally, the use of image based uncertainty for the model evolution during the segmentation process should be studied.

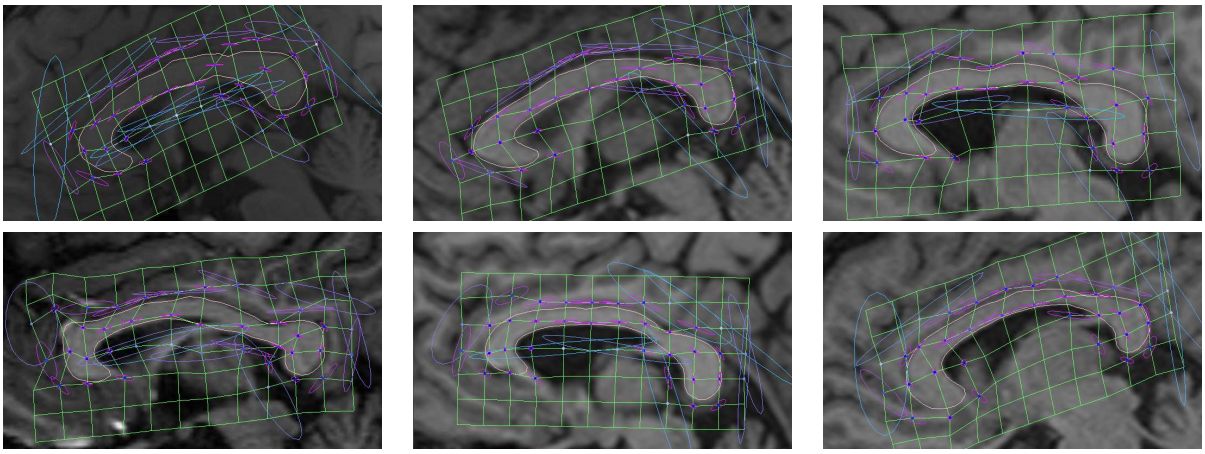


Fig. 4.7: Additional segmentation results with uncertainty measures.

Chapter 5

Conclusions & Future Directions

Abstract – *In this thesis we have investigated shape registration, shape modeling and their applications to knowledge-based segmentation in medical image analysis. The main contribution of our approach consists of introducing the notion of uncertainties in all levels of the process. Such uncertainties aim to account for (i) erroneous segmentation results due to the lack of visual support , (ii) erroneous registration results either due to convergence to local minimum or because geometric correspondences do not always refer to meaningful anatomical correspondences. We have proposed two different statistical models to encode such measurements, one that is organ-based and aim to cope with structures of limited variation which exhibit a parametric nature and a second that is domain-based and does not make any assumptions on the statistical properties of the samples. We have used two different applications to demonstrate the potentials of our approach.*

5.1 Introduction

Despite all the progress made in the area of shape modeling, shape registration, statistical modeling of shape variations and their application to knowledge-based segmentation, one cannot claim the existence of a universal solution to address these problems. However, a strong demand exists to recover such a solution due to the important number of applications which will benefit from it in the domains of medical image analysis, computer vision and computer graphics. While the above-mentioned four tasks are strongly related and there is no doubt that the optimal approach should address jointly the registration and modeling of shape variation components, this is not the case. Most of the existing approaches in the literature decompose the process using a sequential logic. Therefore it is natural to introduce a strong bias in a certain task, if it depends on the results of a previous one. The accumulation of errors in the process will lead to a statistical model with limited justification where an important part of the observable variations are due to the propagation of errors.

This bias can be explained by a number of reasons. First, the training data is not always reliable: in the domain of medical imaging, one can observe an important variability between the results proposed by two different clinical experts. Such a variability increases linearly with the complexity of the task. Then, even if the sample extraction is not erroneous, the task of bringing these samples to the same pose (towards modeling) involves dense registration. This is by definition an ill-posed problem due to the lack of sufficient constraints to recover the solution. Furthermore, while the results can be considered reliable in the presence of geometric features, there is not guarantee that meaningful geometric correspondences are equivalent to proper anatomical ones. Consequently, the use of statistical models to capture the variations of the deformations between the training examples and the common pose will be conditioned by the above errors.

The choice of these models is critical and should be able to capture the variability of the samples as much as possible. Such variability depends on the nature of the organ being considered. The use of parametric models - often considered in the literature - focus mostly on describing this variability without implicitly taking into account the bias due to erroneous extraction of samples or erroneous registration results. These methods strictly focus on the distribution of samples after registration and would produce results that on top of modeling the variation of the samples also encode the error of the registration. One can go even further and claim that the parametric nature of these models imposes a severe constraint on their range and applicability across organs.

One can conclude that there is a great benefit in coping with this bias as well as proposing statistical models of increasing complexity to describe the sample variations. These models should

be as compact as possible while retaining the ability to capture important variations of the sample when these are present. Introducing such an approach was the main motivation of this thesis.

5.2 Contributions

The most natural way to cope with the bias of a sequential process is through the propagation of error from one step to the next. Such an action requires first considering a theoretical model for the sequential process, then recovering qualitative measurements in each step and last propagating these measurements over the process. In order to address this demand, we have considered a registration approach which can associate measurements of uncertainties with the obtained deformation map.

The first major contribution of this thesis consists of proposing two means to determine a qualitative assessment of the obtained registration result. This is done using two different expressions of uncertainties on a shape representation space where there is no explicit need of recovering correspondences. The first was based on the natural Hessian-based expression of the objective function while taking into account properties of the considered feature space. The second was recovered through an explicit modeling of the variations of the data leading to a more general result when compared to the Hessian-based expression. Such an observation has driven us to study the connection between the two different expressions of uncertainties and has produced conditions of mutual understanding. In particular we have shown that the Hessian-based assessment is a particular case of the data-driven one under certain conditions.

Once uncertainties have been determined, the next step consists of propagating them to the statistical model. Such a task involves first the selection of the model and second the use of registrations uncertainties within it. The second major contribution of this thesis consists of introducing a variable bandwidth non-parametric approach to model shape variations. We proposed a definition of the bandwidth that relates to the estimated registration uncertainties and have as aim to decrease the importance of samples or parts of them with limited data support (high uncertainties). This generic model does not make an explicit assumption on the organ to be modeled or on the inter subject variation and can be computationally inefficient. To cope with the above remark, we have also introduced a linear model based on independent component analysis where registration errors are implicitly taken into account through an augmentation of the samples. These models can now be used for segmentation, however it requires an efficient representation of the reference model.

We have proposed two different models of increasing complexity to encode the reference

model. The first one consists of a limited number of control points and an efficient transformation model based on thin plate splines. The second uses a more flexible model which does not make an assumption on the organ and models the deformation space in a uniform fashion using a free from deformation. Two applications, the segmentation of the corpus callosum in mid-sagittal MR-images, and the one of the left ventricle in cardiac CT images have been used to demonstrate the potentials of our approach. Both the transformation models and the knowledge-based segmentation approaches are considered to be the third major contribution of this thesis.

5.3 Future Directions

Despite the introduction of the registration uncertainties on the statistical model the method can still be biased. The first step of the process is the extraction of training samples. Modeling the variation among users as well as associating with the samples measures of confidence based on the observable data support could help us account with this bias. Such uncertainties can be propagated to the registration model and used to define appropriate cost functions where there is coherence between anatomical and geometric correspondences while being able to account for the segmentation bias. The propagation of the segmentation uncertainties along with the ones from registration could produce a statistical model where bias was deduced from the process. Next, better understanding of the registration uncertainties and the relationship between the Hessian and the data-driven model is also to be further investigated. In this thesis we were able to demonstrate some connections between them, however the outcome needs to be further developed.

On the statistical modeling, in terms of parametric modeling we would like to investigate a rather theoretical propagation of the uncertainties. The proposed approach decorelates the uncertainties measurements from the model. Studying models like kernel-based principal component analysis, Kernel-based Independent Component Analysis which can encode the uncertainties measurement directly in the model space is a natural extension of our approach. On the non-parametric modeling approach, the investigation of alternative models to perform dimensionality reduction of the retained samples/kernels is also to be considered. Last, but not least the case of Gaussian Kernels was considered which have some nice implementation properties but limited capture while do not take into account the nature of the application. Therefore, the use of other kernels to describe the deformations of the observed samples and produce a compact density is also to be developed.

On the segmentation component of our approach, we would like first to further develop the notion of qualitative interpretation of the results. Similar to the registration case, we would like to determine cost functions where the use of their Hessian or their data-driven variation can pro-

duce uncertainties on the results. Such a task is challenging in particular when considering the data-driven approach but could have a great impact in the domain of medical image analysis where the availability of clinical users is limited since the method will be able to automatically determine areas where the interaction with the expert would be required. Last, but not least studying the deformations/modeling of more complex organs like the human brain is a very promising direction of our approach for two reasons, (i) these structures exhibit important variability among samples/individuals and therefore the use of non-linear models could have a great impact, and (ii) these structures are characterized by a very fine geometric space and therefore all the steps of the process will provide valuable information towards modeling such a complex system.

Conclusion (Version Française)

Malgré tous les progrès effectués dans le domaine de la modélisation des formes, le recalage de formes, modélisation statistique des variations de formes et segmentation avec a priori de formes, aucune méthode n'a pu s'imposer jusqu'à présent comme une solution universelle. Cependant, il existe une demande importante pour retrouver des solutions à de telles problématiques qui pourront s'appliquer à l'imagerie médicale, la vision par ordinateur. Il apparaît que les quatre problèmes précédemment mentionnés touchent des sujets connexes, une approche optimale devrait donc traiter conjointement le recalage et la modélisation. Pourtant les approches classiques présentées dans l'état de l'art continuent de les traiter indépendamment et de façon purement séquentielle. Par conséquent, il était naturel d'introduire un biais important sur la réalisation d'une certaine tâche si son résultat devra être utilisé pour réaliser une autre tâche. En l'absence de biais, l'accumulation des erreurs à l'étape 1, va créer un modèle statistique difficilement justifiable puisque une partie non négligeable des variations observées sera due à la propagation des erreurs de l'étape 1.

Ce biais est donc une information supplémentaire ayant plusieurs justifications. D'abord, les données utilisées pour l'apprentissage peuvent être imprécises ou partiellement erronées. Ensuite, dans le cas du traitement des images médicales, différents experts fourniront différentes segmentations d'un même organe. Ces variations contribuent donc à augmenter linéairement la complexité du problème. Même si l'extraction des données n'est pas le siège d'erreurs, le fait de transformer tous les échantillons sur le même repère de référence est un problème de recalage dense nécessaire à la modélisation des formes. Ceci reste un problème mal posé dans la mesure où le manque de contrainte dans le recalage ne permet pas de retrouver une solution unique. Par ailleurs, si les résultats obtenus doivent être considérés comme valables dans le cas d'un bon alignement des données géométriques, il n'y a aucune garantie qu'un alignement géométrique corresponde à un alignement des indicateurs anatomiques. Par conséquent, la construction d'un modèle statistique capturant les variations des déformations sera aussi conditionnée par l'accumulation de ces erreurs.

Le choix d'un tel modèle est primordial et doit être capable de représenter au mieux les variations des échantillons, qui dépendent par ailleurs de la nature de l'organe considéré. L'utilisation

d'un modèle paramétrique, fréquemment présenté dans les précédents travaux, se résume trop souvent à décrire les variations sans tenir compte du biais provenant du bruit existant sur les exemples composant les données, ni des erreurs dans l'extraction de ces données à partir des images. Ces méthodes calculent donc une distribution uniquement basée sur le résultat de la phase de recalage et génère donc un modèle pour les variations des échantillons codant aussi les erreurs d'extraction. Par ailleurs, faire le choix d'un modèle paramétrique impose une contrainte importante sur les variations autorisées et la possibilité de les appliquer à différents organes.

L'utilisation de ce biais permet de proposer un modèle statistique décrivant uniquement les variations des échantillons, les erreurs d'extraction font alors partie du modèle. L'introduction d'une telle approche et ces applications sont l'objectif principal de cette thèse.

Synthèse des travaux

L'approche la plus immédiate permettant d'utiliser le biais existant dans une chaîne de processus, nécessite de propager ces informations d'une étape à la suivante. Ce raisonnement est donc construit en considérant d'abord un modèle théorique adapté au processus séquentiel, puis y introduire les mesures de qualité obtenues pour chaque étape et de les utiliser comme autant de données pour l'étape suivante. Dans le cas particulier du recalage et de la modélisation de formes, une approche du recalage a été introduite, ce qui permet de mesurer les incertitudes obtenues sur les déformations paramétriques.

La première contribution de cette thèse propose deux moyens pour mesurer la qualité du résultat obtenu durant la phase de recalage de formes. Nous introduisons pour cela deux expressions pour les incertitudes exprimées dans un espace de représentation des formes tel qu'il n'est pas nécessaire d'avoir une correspondance point à point entre les deux formes. La première expression est basée sur le calcul du Hessien de la fonctionnelle dont le minimum correspond à un recalage géométrique réussi. La seconde expression est basé sur les données, c'est-à-dire les formes elles mêmes sur lesquelles on considère des variations locales aléatoires. Ces variations ayant une influence sur le résultat du recalage, elles permettent de quantifier l'incertitude sur le recalage de chaque forme. Ayant ensuite conduit une étude comparative des deux approches, nous avons pu mettre en évidence d'importants liens aidant à leur compréhension. Nous avons pu montrer en particulier que les deux approches sont identiques lorsque les termes de régularisation du recalage deviennent négligeables.

Dès lors que les incertitudes ont pu être calculées, l'étape suivante permet de les propager

pour la construction du modèle statistique. Cette tâche implique d'abord la sélection d'une forme de référence, mais aussi et surtout la sélection d'un modèle statistique permettant l'utilisation des incertitudes en leur sein. Ceci constitue donc la seconde contribution majeure de ce travail de thèse et introduit une approche non paramétrique utilisant des noyaux de taille et orientation variable pour la modélisation de formes. Nous proposons donc une définition de ces noyaux en relation avec l'évaluation des incertitudes, de sorte que des déformations correspondant à des directions de forte incertitude des noyaux resteront acceptables pour le modèle de forme. Ce modèle générique ne fait aucune supposition préalable sur la nature de l'organe à modéliser. En contrepartie, l'évaluation de la densité estimée pour le modèle de forme devient coûteuse lorsque le nombre d'échantillons augmente. Pour pallier ce problème, nous avons introduit un modèle linéaire basé sur l'analyse en composante indépendante des déformations, dans lequel les incertitudes sont présentes grâce à l'introduction de nouveaux échantillons. Ces modèles statistiques peuvent alors être utilisés pour la segmentation.

Nous avons proposé deux modèles de déformations permettant de transformer la forme de référence. Le premier modèle utilise un nombre limité de points de contrôle et une transformation fondée sur les déformations de plaques minces (TPS). Le second modèle utilise une transformation plus générale, ne nécessitant aucun a priori sur la forme considérée, mais faisant intervenir un plus grand nombre de points de contrôle. Cette transformation fait appel aux déformations de forme libres (FFD). Deux applications ont été étudiées pour montrer la force de cette approche, la segmentation du corps calleux sur des vues sagittales en imagerie à résonance magnétique et la segmentation du ventricule gauche en imagerie scanner.

Perspectives

Malgré l'introduction des incertitudes dans la modélisation statistique, il subsiste un biais non maîtrisé dans notre méthode. Il est dû au fait que les incertitudes sont calculées à partir de variations locales des formes, et ne peut donc tenir compte de grandes déformations. La première étape du processus d'apprentissage étant l'extraction de formes et surfaces à partir d'images, le calcul d'incertitudes pourrait être reconsidéré en tenant compte de ces données visuelles. Ces incertitudes peuvent alors être propagées au modèle statistique et utilisées pour former une fonction de coût apportant cohérence entre les correspondances géométriques et anatomiques alors capables de tenir compte du biais de segmentation. L'utilisation conjointe des incertitudes dues à la segmentation et au recalage permet alors de produire un modèle statistique justifié par l'application finale à la segmentation. Par ailleurs, les liens entre les deux approches de calcul des incertitudes doivent être approfondis, en particulier les liens existant entre les termes de régularisation introduits dans cette

thèse et la régularisation de Tikhonov.

Un autre axe de recherche permet de redéfinir le lien entre la construction des incertitudes et leur utilisation dans le modèle statistique. L'approche actuelle considère construction et utilisation comme étant indépendant. Créer un modèle d'incertitude associé à une mesure particulière dans l'espace des déformations, en termes de géométrie Riemannienne, permettrait alors d'adapter notre approche à des modèles statistiques plus complexes provenant de la théorie des noyaux. On pourrait alors envisager l'analyse en composante principale ou indépendante par noyaux, permettant ainsi d'utiliser des modèles statistiques paramétriques encodant directement les incertitudes. Enfin, l'étude est menée sur des noyaux Gaussien, choisi pour leurs propriétés calculatoires. Cette étude devrait être approfondi à d'autres types de noyaux montrant de meilleures propriétés de convergence dans la théorie de l'apprentissage non paramétrique.

En ce qui concerne la segmentation, le premier axe d'approfondissement touche à l'évaluation des incertitudes sur le résultat obtenu. De façon similaire au recalage de formes, une approche fondée sur le Hessien de la fonctionnelle de segmentation devrait être adoptée et mise en relation avec une approche basée sur les données de l'image. Cette technique aurait alors un impact majeur dans son application au traitement des images médicales. Ceci permettrait de détecter les régions où la segmentation automatique est incertaine, dans lesquelles une interaction du médecin est requise. Enfin, l'étude des déformations et la modélisation d'organes complexes tel que le cerveau humain est un axe de recherche prometteur. Il est en accord avec notre approche non paramétrique pour deux raisons. D'une part ces structures présentent des variations importantes selon les échantillons considérés, favorisant le choix d'un modèle non linéaire. D'autre part ces structures présentent des détails géométriques de faible échelle, de sorte que toutes les étapes de la chaîne présentée dans ces travaux peuvent fournir des informations supplémentaires pour leur modélisation statistique.

Appendix A

Derivation of the Segmentation Image Term

This appendix gives some further exploration of the calculus of the derivative on the energy term E_{image} (equation 4.2). Consider the general formulation of the segmentation problem, where N regions are retrieved, global transformation $\mathcal{A}$ and local parametric deformation $\mathcal{L}(\boldsymbol{\Theta}, .)$ are considered. Consider the inverse transformation of $\mathcal{A} \circ \mathcal{L}$, this diffeomorphism verifies:

$$\mathcal{A}(\mathcal{L}(\boldsymbol{\Theta}, \mathcal{G}(\boldsymbol{\Theta}, \mathbf{y}))) = \mathbf{y} \tag{A.1}$$

Derivation with respect to the parameter $\boldsymbol{\Theta}$ is written:

$$\frac{\partial(\mathcal{A} \circ \mathcal{L})}{\partial \boldsymbol{\Theta}^T}(\boldsymbol{\Theta}, \mathcal{G}(\boldsymbol{\Theta}, \mathbf{y})) + \frac{\partial(\mathcal{A} \circ \mathcal{L})}{\partial \mathbf{x}^T}(\boldsymbol{\Theta}, \mathcal{G}(\boldsymbol{\Theta}, \mathbf{y}))\frac{\partial \mathcal{G}}{\partial \boldsymbol{\Theta}^T}(\boldsymbol{\Theta}, \mathbf{y}) = 0 \tag{A.2}$$

For simpler notation purpose we also pose:

$$D_{\mathcal{C}_i}(\mathbf{x}, \mathbf{y}) = -H(\phi_{\mathcal{R}_{\mathcal{C}_i}}(\mathbf{x}))\log(p_{\mathcal{C}_i}(\mathcal{I}(\mathbf{y})))$$

Then the image term of the energy (eq. 4.2) can be rewritten as:

$$E_{\text{image}}(\boldsymbol{\Theta}) = \int_{\Omega} \sum_{i=1}^{N} D_{\mathcal{C}_i}(\mathcal{G}(\boldsymbol{\Theta}, \mathbf{y}), \mathbf{y}) d\mathbf{y}$$

When differentiating Eq. (A.1) with respect to $\boldsymbol{\Theta}$ and substituting the expression obtained for $d\mathcal{G}/d\boldsymbol{\Theta}$ into the expression of $dE_{\text{image}}(\boldsymbol{\Theta})/d\boldsymbol{\Theta}$, we get the following:

$$\frac{dE_{\text{image}}(\boldsymbol{\Theta})}{d\boldsymbol{\Theta}^T} = -\int_{\Omega} \sum_{i=1}^{N} \frac{\partial D_{\mathcal{C}_i}}{\partial \mathbf{x}^T}(\mathcal{G}(\boldsymbol{\Theta}, \mathbf{y}), \mathbf{y}) \left[\frac{\partial(\mathcal{A} \circ \mathcal{L})}{\partial \mathbf{x}^T}(\mathcal{G}(\boldsymbol{\Theta}, \mathbf{y}), \boldsymbol{\Theta})\right]^{-1} \frac{\partial(\mathcal{A} \circ \mathcal{L})}{\partial \boldsymbol{\Theta}^T}(\mathcal{G}(\boldsymbol{\Theta}, \mathbf{y}), \boldsymbol{\Theta}) d\mathbf{y}$$

Now changing the integration variable according to the diffeomorphism $\mathbf{x} = \mathcal{G}(\mathbf{\Theta}, \mathbf{y})$

$$\frac{dE_{\text{image}}(\mathbf{\Theta})}{d\mathbf{\Theta}} = -\int_{\Omega} \sum_{i=1}^{N} \frac{\partial D_{\mathcal{C}_i}}{\partial \mathbf{x}^T}(\mathbf{x}, \mathcal{A}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x})))\text{com}\left(\frac{\partial(\mathcal{A}\circ\mathcal{L})}{\partial \mathbf{x}^T}(\mathbf{x}, \mathbf{\Theta})\right)^T \frac{\partial(\mathcal{A}\circ\mathcal{L})}{\partial \mathbf{\Theta}^T}(\mathbf{x}, \mathbf{\Theta})d\mathbf{x}$$

where 'com' denotes the matrix of cofactors. When calculating explicitly the partial derivative of D with respect to its first variable, this integral further simplifies into a curve integral along the reference model:

$$\frac{dE_{\text{image}}(\mathbf{\Theta})}{d\mathbf{\Theta}} = -\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \int_{\partial\mathcal{R}_{\mathcal{C}_i}\cap\partial\mathcal{R}_{\mathcal{C}_j}} \tilde{D}_{\mathcal{C}_i/\mathcal{C}_j}(\mathcal{A}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x}))) \left[\text{com}\left(\frac{\partial(\mathcal{A}\circ\mathcal{L})}{\partial \mathbf{x}^T}(\mathbf{x}, \mathbf{\Theta})\right).\nabla\phi_{\mathcal{R}_{\mathcal{C}_i}}(\mathbf{x})\right]^T \frac{\partial(\mathcal{A}\circ\mathcal{L})}{\partial \mathbf{\Theta}^T}(\mathbf{x}, \mathbf{\Theta})d\mathbf{x}$$

with $\tilde{D}$ defined as:

$$\tilde{D}_{\mathcal{C}_i/\mathcal{C}_j}(\mathbf{y}) = -\log(p_{\mathcal{C}_i}(\mathcal{I}(\mathbf{y}))) + \log(p_{\mathcal{C}_j}(\mathcal{I}(\mathbf{y})))$$

This expression of the derivative refers only to the shape boundaries in the space of the model and considers all boundaries defined on the template.

As stated in chapter 2.4, $\phi_{\mathcal{R}_{\mathcal{C}_i}}(\mathcal{G}(\mathbf{\Theta}, \mathbf{y}))$ and $\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R}_{\mathcal{C}_i})}(\mathbf{y})$ have the same 0 iso-level. Therefore the gradient of these two implicit functions are collinear:

$$\begin{aligned}\nabla\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R}_{\mathcal{C}_i})}(\mathbf{y}) &\propto \frac{d(\phi_{\mathcal{R}_{\mathcal{C}_i}}\circ\mathcal{G})}{d\mathbf{y}}(\mathbf{\Theta}, \mathbf{y}) \\ &\propto \text{com}\left(\frac{\partial(\mathcal{A}\circ\mathcal{L})}{\partial \mathbf{x}^T}(\mathcal{G}(\mathbf{\Theta}, \mathbf{y}), \mathbf{\Theta})\right).\nabla\phi_{\mathcal{R}_{\mathcal{C}_i}}(\mathcal{G}(\mathbf{\Theta}, \mathbf{y}))\end{aligned}$$

By considering Heaviside functions defined in the image space (i.e.: $H(\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R}_{\mathcal{C}_i})}(\mathbf{y}))$), then one obtains a simpler expression of the derivative as mentioned in equation (4.3):

$$\frac{dE_{\text{image}}(\mathbf{\Theta})}{d\mathbf{\Theta}^T} = -\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \int_{\partial\mathcal{R}_{\mathcal{C}_i}\cap\partial\mathcal{R}_{\mathcal{C}_j}} \tilde{D}_{\mathcal{C}_i/\mathcal{C}_j}(\mathcal{A}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x}))) \nabla\phi_{\mathcal{L}(\mathbf{\Theta},\mathcal{R}_{\mathcal{C}_i})}(\mathcal{A}\circ\mathcal{L}(\mathbf{\Theta}, \mathbf{x}))^T \frac{\partial(\mathcal{A}\circ\mathcal{L})}{\partial \mathbf{\Theta}^T}(\mathbf{x}, \mathbf{\Theta})d\mathbf{x}$$

Appendix B

Discrete Computation of Distance on Anisotropic Grids

Distance maps were presented in section 2.2.4 as a particular class of images where the value at each pixel refers to the shortest distance to the interface. We will introduce in this section two algorithms known as *Distance transformations*, able to compute distance map on discrete grids either from a binary image or from a consistently oriented mesh. We refer to [50] for an exhaustive review of all exact and approximated distance transform algorithms . We will present two methods able to limit the computation to a narrow band surrounding the zero-level. Narrow band is a widely used technique to speed up level-set evolution, it will also be extensively used within our registration technique based on deformable template. This section begin with a description of discrete gradient required in the fast marching algorithm and in our registration framework.

B.1 Numerical Gradient.

Many numerical schemes were described by Sethian in [164]. In the present work, we simply require the discrete computation of the distance map gradient at any voxel. Defining discrete first order derivatives in 2D:

$$\begin{array}{cc} Centered: D^x_{i,j}f = \frac{f_{i+1,j}-f_{i-1,j}}{2\Delta x} & D^y_{i,j}f = \frac{f_{i,j+1}-f_{i,j-1}}{2\Delta y} \\ Left: D^{+x}_{i,j}f = \frac{f_{i+1,j}-f_{i,j}}{\Delta x} & D^{+y}_{i,j}f = \frac{f_{i,j+1}-f_{i,j}}{\Delta y} \\ Right: D^{-x}_{i,j}f = \frac{f_{i,j}-f_{i-1,j}}{\Delta x} & D^{-y}_{i,j}f = \frac{f_{i,j}-f_{i,j-1}}{\Delta y}, \end{array} \tag{B.1}$$

image gradients can simply be defined using centered derivatives:

$$\vec{\mathbf{n}}_{i,j} = (D^x_{i,j} f, D^y_{i,j} f) \tag{B.2}$$

In the case of approximate distance maps however, the normal vector is better approximated as the mean of non centered derivatives:

$$\tilde{\vec{\mathbf{n}}}_{i,j} = \frac{(D^{+x}_{i,j} f, D^{+y}_{i,j} f)}{\|D^{+x}_{i,j} f, D^{+y}_{i,j} f\|} + \frac{(D^{+x}_{i,j} f, D^{-y}_{i,j} f)}{\|D^{+x}_{i,j} f, D^{-y}_{i,j} f\|} + \frac{(D^{-x}_{i,j} f, D^{+y}_{i,j} f)}{\|D^{-x}_{i,j} f, D^{+y}_{i,j} f\|} + \frac{(D^{-x}_{i,j} f, D^{-y}_{i,j} f)}{\|D^{-x}_{i,j} f, D^{-y}_{i,j} f\|} \tag{B.3}$$

and then normalized: $\vec{\mathbf{n}} = \frac{\tilde{\vec{\mathbf{n}}}}{\|\tilde{\vec{\mathbf{n}}}\|}$

B.2 Fast Marching

Fast Marching is a very widespread method to compute distance from a shape boundary, ie. contour or surface. The basic idea is to consider the motion of the contour or surface along the normal direction with unit speed. Then distance to the initial contour corresponds to the time at which the front crosses a certain location. Formally, this can be written for a parameterized contour $\mathcal{C}(x,t)$ with the motion equation:

$$\begin{aligned} \tfrac{\partial \mathcal{C}(x,t)}{\partial t} &= \mathcal{N} \\ \mathcal{C}(x,0) &= \partial\Omega, \end{aligned} \tag{B.4}$$

where $\mathcal{N}$ represents the normal to the contour.

Let ϕ be the function associating any spatial point with the crossing time of the moving contour, ϕ is equal to the distance transform of the contour and verifies $\phi(x) = 0$ for any point on the shape boundary $\partial\Omega$ and the following eikonal equation:

$$\begin{aligned} \phi(\mathcal{C}(x,t)) \equiv t \quad &\rightarrow \nabla\phi.\mathcal{C}_t(x,t) = 1 \\ &\rightarrow \nabla\phi.\left(\tfrac{\nabla\phi}{|\nabla\phi|}\right) = 1 \\ &\rightarrow |\nabla\phi| = 1 \end{aligned} \tag{B.5}$$

In [165], Sethian proposed a discrete framework based on upwind schemes to solve this equation that preserves the causality of the front propagation and can therefore admit the non-differentiability of the distance on the skeleton. The Eikonal equation can be rewritten on the discrete grid:

$$max(D^{-x}_{i,j}\phi, 0)^2 + min(D^{+x}_{i,j}\phi, 0)^2 + max(D^{-y}_{i,j}\phi, 0)^2 + min(D^{+y}_{i,j}\phi, 0)^2 = 1 \tag{B.6}$$

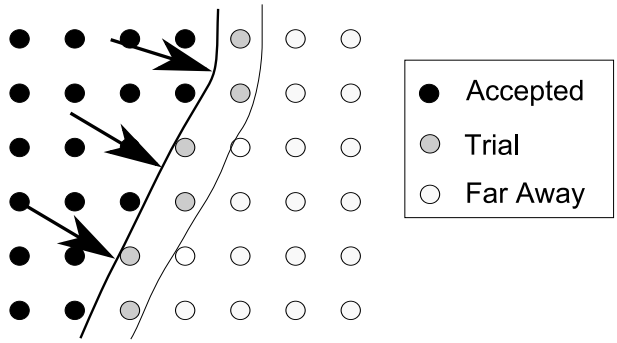


Fig. B.1: The 3 types of points in Fast Marching algorithm

In [156], Rouy and Tourin proposed to solve this equation in an iterative manner directly on the whole grid. The computational cost was too expensive and the front evolution method shows a more efficient framework. The image domain is divided into 3 regions, defining 3 different types of points with an associated distance value (see Fig. B.1):

- *Accepted*: These points are located behind the front. Crossing time is known and their value will not change.
- *Far Away*: These points are located after the front which has not reach these locations yet. Crossing time is arbitrarily fixed to $+\infty$ on these pixels.
- *Trial*: The front is presently standing on these points, their crossing time is estimated using *Accepted* neighboring points. Their value may change as long as their status do not change to *Accepted*.

Initialization of the distance value will be discussed later in the appendix. Assume for now that we dispose of a set $P1$ of *Accepted* points where the value is known. Every neighbor points of $P1$ in a 4-connected neighborhood is set to *Trial* and their value is computed based on equation B.6 using uniquely *Accepted* points values. All other points are set to *Far Away* and their value is set to $+\infty$.

The algorithm is then made of two steps repeating until the smallest trial point value is above the narrow band size:

1. Select the *Trial* point $P0$ with the smallest value, set this point state to *Accepted*.
2. Set every non-accepted neighbor point of $P0$ to *trial* and compute its value using equation B.6 with the newly accepted point.

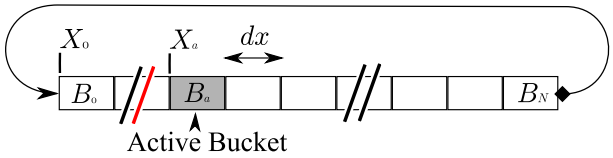


Fig. B.2: The circular list of [211], containing N buckets with size dx. The active buckets Ba contains the *trial* points that are removed from the list and set to *accepted*.

The main difficulty of the algorithm is to find the smallest *Trial* point. This is performed using min-heap sorting and constantly keeping the trial points in a sorted list. The cost of insertion of a new point value in this list is $O(\log n)$ so that fast marching is commonly known as a $O(n.\log n)$ method.

Recently in [211], an $O(n)$ implementation was proposed, based on a discussion of fast marching precision, the authors proposed to remove the min-heap sort and use a circular list instead. The list is made of a number of buckets, each containing a set of trial points with value located within a range dx. The point removed from *Trial* and set to *Accepted* may no longer be the smallest one but belongs to the bucket with the smallest values (see Fig. B.2). Newly updated *Trial* neighbors are added to the bucket corresponding to their value: $(\lfloor (v - x0)/dx \rfloor + B0) \mod N$. It is necessary that the number of buckets in the list is sufficient to enforce the condition that all values in a bucket lies within the range dx, that is equivalent to $(v - xA) < N * dx$. This is easily enforced in the case of distance transform as the difference in value between neighboring points is bounded and cannot exceed the pixel size.

B.3 Chamfer Distance

Different to fast marching, chamfer distance do not require to handle lists of points and is always computed in linear time. In order to recall the principle of Chamfer distance transform, we need to introduce some definitions. The image domain I is associated with a finite subset of $\mathbb{Z}^2$. Therefore we can talk about basis in the domain I for the sum defined on $\mathbb{Z}^2$ and introduce the definition of chamfer mask:

$$\mathcal{M}_C = \{(\mathbf{v}_k, \omega_k) \in \mathbb{Z}^2 \times \mathbb{R}, 1 \leq k \leq 2\},$$

which do not contain the 0 vector, contains a basis of $\mathbb{Z}^2$, has positive weights ($\omega_k > 0$) and verifies all central and axial symmetries ($(\mathbf{v}, \omega) \in \mathcal{M}_C \Rightarrow ((\pm v_i), \omega) \in \mathcal{M}_C$ where (v_i) are the coordinates of $\mathbf{v}$). Chamfer mask contains displacement vectors on the image grid with an associated weight. As it contains a basis or $\mathbb{Z}^2$ it allows to reach any two points of the image with a finite sum of

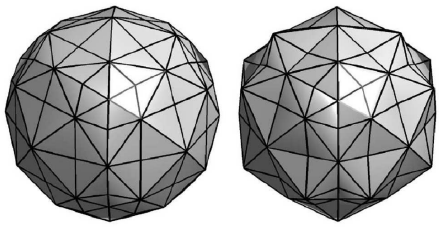


Fig. B.3: The rational balls in 3D for a $(5 \times 5 \times 5)$. Left : the chamfer mask induce a norm on $\mathbb{Z}^3$, Left : the mask do not induce a norm on $\mathbb{Z}^3$ (from [67])

vectors in $\mathcal{M}_C$:

$$\vec{AB} = \sum_{\mathbf{v}_k \in \mathcal{M}_C} n_k.\mathbf{v}_k.$$

This sum also defines a path on the image between A and B and its associated cost is:

$$W(\mathcal{P}_{AB}) = \sum_{k=1}^{m} n_k.\omega_k.$$

The chamfer distance between A and B is simply defined as the minimum cost for all paths between A and B: $d_{AB} = minW(\mathcal{P}_{AB})$. The rational chamfer mask is defined from the chamfer mask as the division of every vector by its associated weight in $\mathbb{R}^2$:

$$\mathcal{M'}_C = \{(\mathbf{v}_k/\omega_k) \in \mathbb{R}^2, 1 \leq k \leq 2\},$$

Then, it can be shown that the chamfer mask defines a distance on $\mathbb{Z}^2$ if and only if the triangulation of the associated rational mask defines a convex solid (see Fig. B.3).

The computation of optimal chamfer coefficients as the best approximation of the Euclidean distance has been introduced for a $2D$, 3×3 mask in [16], and then extended and discussed for higher dimensions and larger masks [201, 111]. In [67, 107], an in depth discussion has been carried on the general choice of optimal integer coefficients on the mask.

In our case the distance were computed on the anisotropic grids by using a $3 \times 3 \times 3$ mask where the value of the coefficients were real and equal to the euclidean distance between the neighboring nodes of the grid.

From a computational point of view, Chamfer distance is computed in two sequential operations using symmetric half Chamfer masks [151]. The image is initialized with 0 in the background and ∞ on the object. Any Chamfer mask $\mathcal{M}_C$ is decomposed into two symmetric mask: a forward mask $\mathcal{M}_C^f$ associated to a natural scan of the image (left to right and top to bottom), and a back-

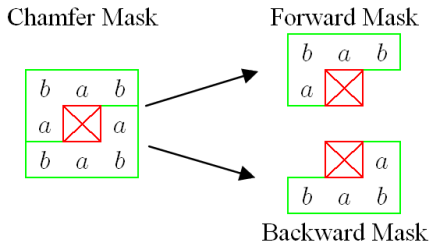


Fig. B.4: Decomposition of a Chamfer mask, forward and backward on a 3×3 mask

ward mask associated to a reverse scan of the image (right to left and bottom to top) as shown on Fig B.4.

The process on the whole image is replacing the current value of a pixel $(I(p))$ according to the rule:

$$I(p) = \min(I(p), \min_{\mathbf{v}_k \int \mathcal{M}_C^f} (I(p + \mathbf{v}_k) + \omega_k),$$

Using a similar process with the backwards mask, this leads to the chamfer distance transform as detailed in figure B.5 with a 3×3 grid. Generalization to 3D is straightforward.

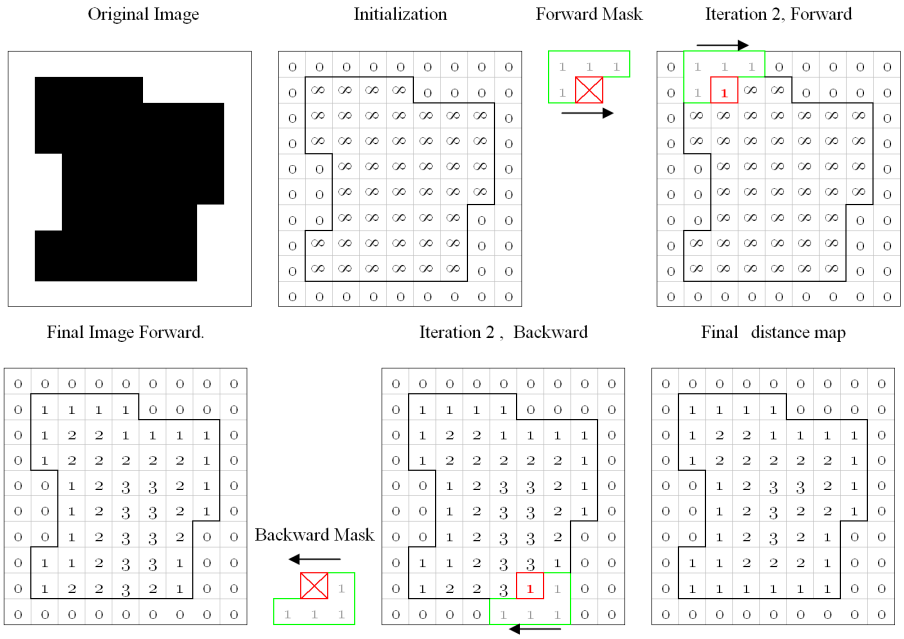


Fig. B.5: Sequential computation of the chamfer distance

B.4 Signed Distance Initialization

In order to represent shapes with signed distance a proper initialization of the distance map is necessary. Both methods presented above assume the existence are propagating pixels based on the existence of a set of pixels with known value.

A classical approach used upon reinitialization of the level set framework consists in isolating the zero crossing of the level set with a very thin band (2 pixels) and recomputing the distance function with some iterations produced with an explicit scheme on the PDE:

$$\frac{\partial \phi}{\partial t} = sign(\phi).(1 - |\nabla \phi|)$$

In the present case of initialization from a binary shape, we use a direct approach to compute the distance to the pixel boundary considering every pixels as a squared dot. This is performed using the chamfer mask presented in the previous section where every weight ω_i is divided by two as shown on figure B.6. This method is particularly adapted to chamfer masks with size $3 \times 3 \times 3$.

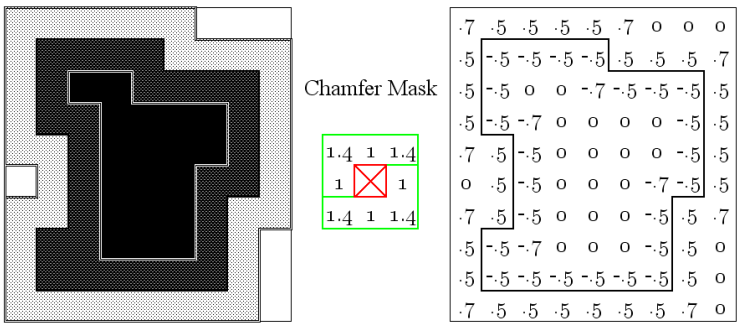


Fig. B.6: Initialisation of distance from binary images

Some other applications require the initialization of a distance map from a discretized contour in 2D (resp. surface in 3D). In such case signed distance computation requires the contour (resp. surface) to be locally oriented with a unit normal vector at every segments (resp. face). Then, every pixel located in the bounding box of every segment (Resp. face) is initialized to the actual Euclidean distance, signed according to the orientation B.7. This initialization requires to pay a particular attention to pixels whose closest point is located at an edge of the contour (resp. edge and vertex of the surface). For such points the sign of the Euclidean distance is set according to the edge normal vector, average of the neighboring segments normal vectors as shown on figure B.7. In the 3 dimensional case, initialization from a triangulated mesh is performed in a similar

fashion with the method presented in [7], where the sign of the distance function around edges and vertices is computed using angle weighted pseudo-normal.

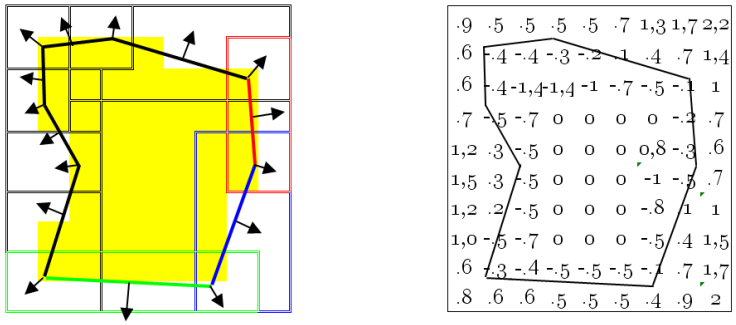


Fig. B.7: Initialisation of distance from an oriented contour

Appendix C

Uncertainty Driven, Kernel Based Shape Modeling for Digits Recognition

In order to evaluate the performance of the method, regarding both the registration and the modeling part, we considered an application with an large training presenting large variability among shapes. Character and in particular digit recognition is a problem that meets these requirements as it shows very large variations among individual examples.

Since the aim of this application was to provide a proof performance of the method rather than a complete recognition system, we focused on three digits which bear certain similarities, namely the '3', '4' and '9'. For each class, we have used 2000 examples of each digit from the MNIST digit database [98] to build the prior model. We then used the kernel selection algorithm 3.6 to retain 50 kernels.

The initial step consists of building an 'average' model for each of the 3 digits. To do so we refer to the classical method presented in [46]. Although an average model is not required within our non parametric density estimator, it is desirable to present a smooth shape that has the ability of being registered to all samples of the training set without topology change.

It should be noted at this stage that the process used for digit recognition is exactly the same as the one followed on learning: the digit model is initially registered to the target with respect to an affine transformation, then local FFD registration is performed after subtraction of the global transformation with uncertainty computed upon registration as presented in section (2.5). This uncertainty estimate is used within the hybrid density estimator presented in section (3.5) to get the final matching score.

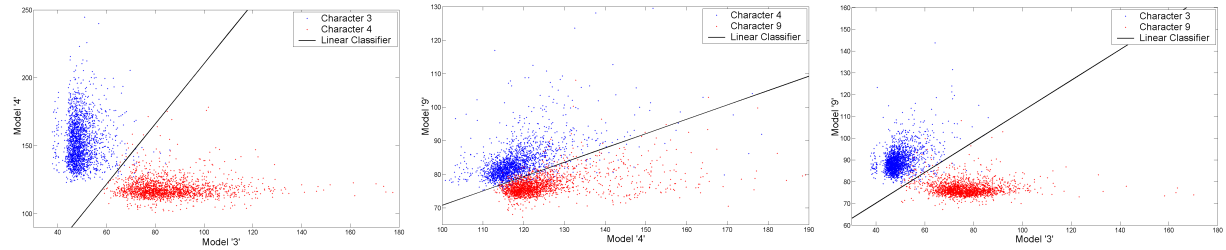


Fig. C.1: (left) Distribution of the digits 3 and 4 in the space of likelihoods of belonging to the classes '3' and '4'. (middle) Distribution of the digits 4 and 9 in the space of likelihoods of belonging to the classes '4' and '9'. (right) Distribution of the digits 3 and 9 in the space of likelihoods of belonging to the classes '3' and '9'.

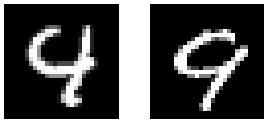


Fig. C.2: Digits 4 and 9 can be very similar.

To verify that our method can encode the shape properties of the class of objects of interest, we ran a cross validation test, where each of the 3 models was registered to all 6000 digits. We then computed the hybrid estimator for the probability of the digit belonging to the class of the model. Fig. C.1 shows the results. Fig. C.1a represents the matching of '3' and '4'. The X-axis is the likelihood that an example belongs to the class of '3' ($-log(\hat{f}(x, \Sigma))$) and the Y-axis is the likelihood that an example belongs to the class '4'. It can be seen that the two classes are very well separated. To demonstrate the separation, we used a simple support vector machine classifier [104] to linearly separate the two classes in the space of likelihood measured. The linear boundary is also shown in Fig. C.1a. The correct classification rate was 99.17%. Fig. C.1b illustrates the separation between classes '3' and '9', the correct classification rate was 98.73%. Finally, Fig. C.1c illustrates the separation between classes '4' and '9', the correct classification rate was 94.83%. Table C.1 shows the overall confusion matrix. The results are consistent with what was expected. The lowest classification rate was obtained when comparing the '4' and the '9'. These digits are indeed very similar when handwritten by Americans, as can be seen from Fig. C.1. We can also see that '3' and '9' look more alike than '3' and '4'. It is important to note that the proposed method is not intended for such an application. However, given this validation we claim that such a model can capture samples of increasing complexity. Also, the use of deformations along with uncertainties provide efficient density estimators.

In order to demonstrate the performance of the variable-bandwidth density approximation, we have compared our method with the case of fixed-bandwidth isotropic Gaussian Kernels. For

the later case, the median value of the Euclidan distances between each samples and its nearest neighbor in the training set was considered. Such a bandwidth is a common choice in the domain of non parametric density estimation. We also applied the reduction algorithm presented in subsection (3.6) to limit the complexity of the method. The results of the comparison are shown in Table C.1. We can observe that for the case of false-negative (recognition error) an improvement that varies between three times (notice the case of '4' - fixed bandwidth error: 20%, variable bandwidth: 7%) to seven times (case of '9', - fixed bandwidth error: 27%, variable bandwidth: 6%) is obtained from the use of variable bandwidth approximation. On the other hand, the error (false-positive) of fixed-bandwidth is in most of the cases 5 times more important than the one obtained from our method. Such an improvement is due to the orientation of bandwidth, capable of discrimating registration results leading to similar transformations. This situation is not so rare when considering '4' and '9' and explain the superiority of our method.

	'3'	'4'	'9'
'3'	0.9240	0.0435	0.0325
'4'	0.0270	0.8035	0.1695
'9'	0.0530	0.2105	0.7365

	'3'	'4'	'9'
'3'	0.9845	0.0065	0.0090
'4'	0.0045	0.9385	0.0570
'9'	0.0145	0.0425	0.9430

Tab. C.1: Confusion matrix between the three classes of digits '3', '4', and '9'. Case of constant isotropic Gaussian kernels on the left, our Method on the right

Bibliography

[1] CGAL, Computational Geometry Algorithms Library. http://www.cgal.org.

[2] I. Abramson. Arbitrariness of the pilot estimator in adaptive kernel methods. *J. Multivariate Anal.*, 12:562–567, 1982.

[3] S. Allassonnière, Y. Amit, and A. Trouvé. Toward a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society*, 69:3–29, 2007.

[4] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, NY, second edition, 1984.

[5] H. Asada and M. Brady. The curvature primal sketch. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):2–14, 1986.

[6] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson. Image segmentation using active contours: calculus of variations or shape gradient? *Siam Journal on Applied Mathematics*, 63(6):2128–2154, 2003.

[7] J. Bærentzen and H. Aanæs. Signed distance computation using the angle weighted pseudonormal. *IEEE Transactions on Visualization and Computer Graphics*, 11(3):243–253, may 2005.

[8] E. Bardinet, L. D. Cohen, and N. Ayache. Tracking medical 3D data with a deformable parametric model. In *ECCV*, pages 317–328, 1996.

[9] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[10] S. Belongie, J. Malik, and J. Puzicha. Matching Shapes. In *IEEE International Conference in Computer Vision*, pages 456–461, 2001.

[11] S. Berretti, A. del Bimbo, and P. Pala. Retrieval by shape using multidimensional indexing structures. In *ICIAP '99: Proceedings of the 10th International Conference on Image Analysis and Processing*, page 945, Washington, DC, USA, 1999. IEEE Computer Society.

[12] J. Besag. On the statistical analysis of dirty images. *Journal of Royal Statistics Society*, 48:259–302, 1986.

[13] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[14] H. Blum. A Transformation for Extracting New Descriptors of Shape. In W. Wathen-Dunn, editor, *Models for the Perception of Speech and Visual Form*, pages 362–380. MIT Press, Cambridge, 1967.

[15] F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[16] G. Borgefors. Distance Transformations in Digital Images. *Computer Vision, Graphics and Image Processing*, 34:344–371, 1986.

[17] J. Bosch, S. Mitchell, B. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. Reiber. Automatocic Segmentation of Echocardiographic Sequences by Active Appearance Motion Models. *IEEE Transactions on Medical Imaging*, 21:1374–1383, 2002.

[18] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1124–1137, 2004.

[19] L. Breiman, W. Meisel, and E. Purcell. Variable kernel estimates of multivariate densities. *Technometrics*, 19:135–144, 1977.

[20] O. Camara, G. Delso, and I. Bloch. Free form deformations guided by gradient vector flow: A surface registration method in thoracic and abdominal pet-ct applications. In *WBIR*, pages 224–233, 2003.

[21] A. Can, C. Stewart, B. Roysam, and H. Tanenbaum. A Feature-Based, Robust, Hierarchical Algorithm for Registering Pairs of Images of the Curved Human Retina. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:347–364, 2002.

[22] J. Canny. A computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:769–798, 1986.

[23] J.-F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.

[24] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. In *IEEE International Conference in Computer Vision*, pages 694–699, 1995.

[25] T. Chan and L. Vese. An Active Contour Model without Edges. In *International Conference on Scale-Space Theories in Computer Vision*, pages 141–151, 1999.

[26] T. Chan and L. Vese. Active Contours without Edges. *IEEE Transactions on Image Processing*, 10:266–277, 2001.

[27] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Found. Comput. Math.*, 5(1):1–58, 2005.

[28] G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, and O. Faugeras. Generalized gradients: Priors on minimization flows. *International Journal of Computer Vision*, 73(3):325–344, 2007.

[29] H. Chen and P. Meer. Robust computer vision through kernel density estimation. In *ECCV (1)*, pages 236–250, 2002.

[30] Y. Chen and G. Medioni. Object Modeling by Registration of Multiple Range Images. *Image and Vision Computing*, 10:145–155, 1992.

[31] Y. Chen, H. Thiruvenkadam, H. Tagare, F. Huang, and D. Wilson. On the Incorporation of Shape Priors int Geometric Active Contours. In *IEEE Workshop in Variational and Level Set Methods*, pages 145–152, 2001.

[32] Y. Cheng. Mean Shift, Mode Seeking, and Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:790–799, 1995.

[33] H. Chui and A. Rangarajan. A New Algorithm for Non-Rigid Point Matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages II: 44–51, 2000.

[34] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, 2003.

[35] I. Cohen and L. D. Cohen. A hybrid hyperquadric model for 2-d and 3-d data fitting. *Computer Vision and Image Understanding*, 63(3):527–541, 1996.

[36] L. Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, 53:211–218, 1991.

[37] L. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, November 1993.

[38] L. D. Cohen. On active contour models and balloons. *Computer Vision, Graphics, and Image Processing*, 53(2):211–218, 1991.

[39] D. Comaniciu and P. Meer. Mean shift analysis and applications. In *IEEE International Conference in Computer Vision*, page 1197, Washington, DC, USA, 1999. IEEE Computer Society.

[40] D. Comaniciu and P. Meer. Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619, 2002.

[41] D. Comaniciu, V. Ramesh, and P. Meer. The variable bandwidth mean shift and data-driven scale selection. In *IEEE 8th Int. Conf. on Computer Vision*, pages 438– 445, 2001.

[42] T. Cootes, C. Beeston, G. Edwards, and C. Taylor. Unified Framework for Atlas Matching Using Active Appearance Models. In *Information Processing in Medical Imaging*, pages 322–333, 1999.

[43] T. Cootes, A. Hill, C. Taylor, and J. Haslam. Use of Active Shape Models for Locating Structures in Medical Imaging. *Image Vision and Computing*, 12:355–366, 1994.

[44] T. Cootes and C. Taylor. Mixture model for representing shape variation. *Image and Vision Computing*, 17:567–574, 1999.

[45] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Training models of shape from sets of examples. In *British Machine Vision Conference*, pages 9–18. Springer-Verlag, 1992.

[46] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.

[47] D. Cremers, T. Kohlberger, and C. Schnörr. Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36:1929–1943, 2003.

[48] D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International Journal of Computer Vision*, 69(3):335–351, 2006.

[49] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion snake : Introducing statistical shape knowledge into the mumford-shah functional. *International Journal of Computer Vision*, 3:295–313, 2002.

[50] O. Cuisenaire. *Distance transformations: fast algorithm and applications to medical image processing*. PhD thesis, Universite´ Catholique de Louvain, October 1999.

[51] E. Davies. *Machine Vision: Theory, Algorithms, Practicalities*, page 171–191. Academic Press, New York, 1997.

[52] R. H. Davies, T. F. Cootes, and C. J. Taylor. A minimum description length approach to statistical shape modelling. In *Information Processing in Medical Imaging*, pages 50–63, London, UK, 2001. Springer-Verlag.

[53] C. de Boor. *A practical guide to splines*. New York: Springer-Verlag, 1978.

[54] M. Delfour and J.-P. Zolésio. *Shapes and geometries : analysis, differential calculus, and optimization*. Siam, 2001.

[55] H. Delingette. General object reconstruction based on simplex meshes. *International Journal of Computer Vision*, 32(2):111–146, September 1999.

[56] A. Dervieux and F. Thomasset. A finite element method for the simulation of rayleigh-taylor instability. *Lect. Notes Math.*, 771:145–159, 1979.

[57] A. Dervieux and F. Thomasset. Multifluid incompressible flows by a finite element method. In R. M. W. Reynolds, editor, *Seventh Int. Conf. on Numerical Methods in Fluid Dynamics*, volume 141 of *Lect. Notes Phys*, pages 158–163, 1980.

[58] G. Donato and S. Belongie. Approximate thin plate spline mappings. In *European Conference on Computer Vision*, pages 21–31, 2002.

[59] R. Donner, M. Reiter, G. Langs, P. Peloschek, and H. Bischof. Fast active appearance model search using canonical correlation analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1690–1694, Oct 2006.

[60] D. Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52:1289 – 1306, April 2006.

[61] Y. Duan and H. Qin. A subdivision-based deformable model for surface reconstruction of unknown topology. *Graph. Models*, 66(4):181–202, 2004.

[62] J. Duchon. Spline minimizing rotation invariant semi norms in sobolev spaces. In W. Schempp and K. Zeller, editors, *Constructive theory of functions os several variables*. Springer Verlag, 1977.

[63] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.

[64] P. Etyngier, F. Ségonne, and R. Keriven. Shape priors using manifold learning techniques. In *IEEE International Conference in Computer Vision*, Rio de Janeiro, Brazil, Oct 2007.

[65] M. Figueiredo, J. M. N. Leitao, and A. K. Jain. On fitting mixture models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, page 54–69, 1999.

[66] C. Florin, N. Paragios, G. Funka-Lea, and J. Williams. Liver segmentation using sparse 3d prior models with optimal data support. In *Information Processing in Medical Imaging*, pages 38–49, 2007.

[67] C. Fouard and G. Malandain. 3-d chamfer distances and norms in anisotropic grids. *Image and Vision Computing*, 23(2):143–158, February 2005.

[68] A. Foulonneau, P. Charbonnier, and F. Heitz. Affine-invariant geometric shape priors for region-based active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1352–1357, 2006.

[69] C. Fraley and A. Raftery. How many clusters? which clustering method? answers via model-based cluster analysis. Technical Report 329, Department of Statistics, University of WashingtonUniversity of Washington, Seattle, WA, 1998.

[70] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[71] F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural networks architectures. *Neural Computation*, 7(2):219–269, 1995.

[72] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 1989.

[73] J. Gomes and O. Faugeras. Reconciling distance functions and level sets. *Journal of Visual Communication and Image Representation*, 11:209–223, 2000.

[74] L. Grady, Y. Sun, and J. Williams. Three interactive graph-based segmentation methods applied to cardiovascular imaging. In O. F. Nikos Paragios, Yunmei Chen, editor, *Handbook of Mathematical Models in Computer Vision*, pages 453–469. Springer, 2006.

[75] U. Grenander. *General Pattern Theory*. Oxford University Press, 1993.

[76] A. Guéziec and N. Ayache. Smoothing and matching of 3-D space curves. *International Journal of Computer Vision*, 12(1):79–104, January 1994.

[77] H. Gävert, J. Hurri, J. Särelä, and A. Hyvärinen. Fastica - fast independent component analysis, 2005.

[78] P. Hall, T. Hui, and J. Marron. Improved variable window kernel estimates of probability densities. *The Annals of Statistics*, 23(1):1–10, 1995.

[79] R. Haralick. Digital step edges from zero crossing of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:58–68, 1984.

[80] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[81] B. Horn and B. Schunck. Determinating Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.

[82] X. Huang, D. Metaxas, and T. Chen. Metamorphs: Deformable shape and texture models. In *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, CVPR'04*, volume I, pages 496–503, 2004.

[83] X. Huang, N. Paragios, and D. Metaxas. Registration of Structures in Arbitrary Dimensions: Implicit Representations, Mutual Information & Free-Form Deformations. Technical Report DCS-TR-0520, Division of Computer & Information Science, Rutgers University, 2003.

[84] X. Huang, N. Paragios, and D. N. Metaxas. Shape registration in implicit spaces using information theory and free form deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1303–1318, 2006.

[85] P. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[86] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[87] A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13:411–430, 2000.

[88] M. Isard and A. Blake. Contour Tracking by Stochastic Propagation of Conditional Density. In *European Conference on Computer Vision*, volume I, pages 343–356, 1996.

[89] J. Jackson. *A User's Guide to Practical Principal Components*. Wiley, 1991.

[90] O. Juan, R. Keriven, and G. Postelnicu. Stochastic motion and the level set method in computer vision: stochastic active contours. *International Journal of Computer Vision*, 69(1):7–25, 2006.

[91] K. Kanatani. Uncertainty modeling and model selection for geometric inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1307–1319, 2004.

[92] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. In *IEEE International Conference in Computer Vision*, pages 261–268, 1987.

[93] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1:321–332, 1988.

[94] D. G. Kendall. The diffusion of shape. *Advances in Applied Probability*, 9:428–430, 1977.

[95] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *IEEE International Conference in Computer Vision*, pages 810–815, 1995.

[96] A. Klassen, E.; Srivastava. Geometric analysis of planar shapes using geodesic paths. In *Signals, Systems and Computers*, volume 1, pages 468 – 472, 2002.

[97] M. Kortgen, G. J. Park, M. Novotni, and R. Klein. 3d shape matching with 3d shape contexts. *In the 7th Central European Seminar on Computer Graphics*, April 2003.

[98] Y. LeCun. The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist/.

[99] M. Leventon, E. Grimson, and O. Faugeras. Statistical Shape Influence in Geodesic Active Contours. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 316–322, 2000.

[100] F. W. B. Li and R. W. H. Lau. Real-time rendering of deformable parametric free-form surfaces. In *VRST '99: Proceedings of the ACM symposium on Virtual reality software and technology*, pages 131–138, New York, NY, USA, 1999. ACM Press.

[101] S. X. Liao and M. Pawlak. On image analysis by moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):254–266, 1996.

[102] E. Llorens, D.; Vidal. Application of extended generalized linear discriminant functions (egldf) to optical character recognition. In *Handwriting Analysis and Recognition - A European Perspective, IEE Workshop on*, volume 23, pages 701–714, 1996.

[103] A. Lundervold, N. Duta, T. Taxt, and A. Jain. Model-guided segmentation of corpus callosum in MR images. In *CVPR*, pages 1231–1238, 1999.

[104] J. Ma, Y. Zhao, and S. Ahalt. OSU SVM classifier matlab toolbox. http://www.ece.osu.edu/∼maj/osu_svm/.

[105] J. MacQueen. Some methods for classification and analysis of multivariate. In *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, page 281–297, 1967.

[106] W. R. Madych and S. A. Nelson. Polyharmonic cardinal splines: a minimization property. *J. Approx. Theory*, 63(3):303–320, 1990.

[107] G. Malandain and C. Fouard. On optimal chamfer masks and coefficients. Research Report 5566, INRIA, Sophia Antipolis, May 2005.

[108] R. Malladi, J. Sethian, and B. Vemuri. Shape Modeling with Front Propagation: A Level Set Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:158–175, 1995.

[109] S. Manay, D. Cremers, B.-W. Hong, A. J. Yezzi, and S. Soatto. Integral invariants for shape matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1602–1618, 2006.

[110] S. Manay, B.-W. Hong, A. J. Yezzi, and S. Soatto. Integral invariant signatures. In *ECCV (4)*, pages 87–99, 2004.

[111] J.-F. Mangin, I. Bloch, J. Lopez-Krahe, and V. Frouin. Chamfer Distances in Anisotropic 3D Images. In *EUSIPCO 94*, pages 975–978, Edinburgh, UK, Sept. 1994.

[112] K. Mardia and I. Dryden. Shape distributions for landmark data. *Advances in Applied Probability*, 4:742–755, 1989.

[113] J. S. Marron and D. Nolan. Canonical kernels for density estimation. *Statistics & Probability Letters*, 7(3):195–199, 1988.

[114] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[115] T. McInerney and D. Terzopoulos. A Dynamic Finite Element Surface Model for Segmentation and Tracking in Multidimensional Medical Images with Application to Cardiac 4D Image Analysis. *Computerized Medical Imaging and Graphics*, 19:69–83, 1995.

[116] G. McLachlan, D. Peel, and W. Whiten. Maximum likelihood clustering via normal mixture model. *Signal Processing: Image Communication*, 8:105–111, 1996.

[117] G. Medioni, S. Menet, and P. Saint-Marc. Active contour models: Overview, implementation and application. In *International Conference on Systems, Man, and Cybernetics*, pages 194–199. IEEE, November 1990.

[118] M. Mignotte, J. Meunier, and J.-C. Tardif. Endocardial boundary e timation and tracking in echocardiographic images using deformable template and markov random fields. *Pattern Anal. Appl.*, 4(4):256–271, 2001.

[119] S. Mika, B. Schölkopf, A. J. Smola, K.-R. Müller, M. Scholz, and G. Rätsch. Kernel PCA and de–noising in feature spaces. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems 11*. MIT Press, 1999.

[120] J. Montagnat and H. Delingette. 4D deformable models with temporal constraints : application to 4D cardiac image segmentation. *Medical Image Analysis*, 9(1):87–100, February 2005.

[121] J. Montagnat, H. Delingette, and N. Ayache. A review of deformable surfaces: topology, geometry and deformation. *Image and Vision Computing (IVC)*, 19(14):1023–1040, Dec. 2001.

[122] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 22–26, 1985.

[123] D. Mumford and J. Shah. Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics*, 42:577–685, 1989.

[124] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[125] C. Nastar and N. Ayache. Frequency-based non-rigid motion analysis: Application to four dimensional medical images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11):1067–1079, novembre 1996.

[126] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer Verlag, 2003.

[127] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed : Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[128] N. Paragios. *Geodesic Active Regions and Level Set Methods: Contributions and Applications in Artificial Vision*. PhD thesis, I.N.R.I.A./University of Nice-Sophia Antipolis, 2000. http://www.inria.fr/RRRT/TU-0636.html.

[129] N. Paragios, Y. Chen, and O. Faugeras, editors. *Handbook of Mathematical Models in Computer Vision*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[130] N. Paragios and R. Deriche. Geodesic Active Regions for Motion Estimation and Tracking. In *IEEE International Conference in Computer Vision*, pages 688–674, 1999.

[131] N. Paragios and R. Deriche. Geodesic Active Regions for Supervised Texture Segmentation. In *IEEE International Conference in Computer Vision*, pages 926–932, 1999. Previous: INRIA Research Report, RR 3440, June 1998, http://www.inria.fr/RRRT/RR-3440.html.

[132] N. Paragios and R. Deriche. Geodesic Active Regions: A New Framework to Deal with Frame Partition Problems in Computer Vision. *Journal of Visual Communication and Image Representation*, 13:249–268, 2002.

[133] N. Paragios and R. Deriche. Geodesic Active Regions Level Set Methods for Supervised Texture Segmentation. *International Journal of Computer Vision*, 46(3):223–247, 2002.

[134] N. Paragios, M. Rousson, and V. Ramesh. Matching Distance Functions: A Shape-to-Area Variational Approach for Global-to-Local Registration. In *European Conference on Computer Vision*, pages II:775–790, 2002.

[135] N. Paragios, M. Rousson, and V. Ramesh. Non-Rigid Registration Using Distance Functions. *Computer Vision and Image Understanding*, 89(2-3):142–165, 2003.

[136] N. Paragios, M. Taron, X. Huang, M. Rousson, and D. Metaxas. On the represantation of shapes using implicit functions. In K. . Yezzi, editor, *Statistics and Analysis of Shapes*, pages 167–200. Springer Verlag, 2006.

[137] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33:1065–1076, 1962.

[138] X. Pennec. Multiple registration and mean rigid shape - application to the 3D case. In K. Mardia, C. Gill, and D. I.L., editors, *Image Fusion and Shape Variability Techniques*

(16th Leeds Annual Statistical Workshop), pages 178–185. University of Leeds, UK, July 1996.

[139] X. Pennec and J.-P. Thirion. A framework for uncertainty and validation of 3-d registration methods based on points and frames. *International Journal of Computer Vision*, 25(3):203–229, 1997.

[140] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):715–729, 1991.

[141] M. Peura and J. Iivarinen. Efficiency of simple shape descriptors. In *Proceedings of the Third International Workshop on visual Form*, pages 443–451, May 1997.

[142] Q. Pham, F. Vincent, P. Clarysse, P. Croisille, and I. Magnin. A fem-based deformable model for the 3d segmentation and tracking of the heart in cardiac mri. In *Image and Signal Processing and Analysis*, 2001.

[143] S. M. Pizer, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash, J. M. Glotzer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, and E. L. Chaney. Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision*, 55(2-3):85–106, 2003.

[144] J.-P. Pons and J.-D. Boissonnat. Delaunay deformable models: Topology-adaptive meshes based on the restricted delaunay triangulation. In IEEE, editor, *Computer Vision and Pattern Recognition*. IEEE, June 2007.

[145] X. Provot. Deformation constraints in a mass-spring model to describe rigid cloth behavior. In W. A. Davis and P. Prusinkiewicz, editors, *Graphics Interface '95*, pages 147–154. Canadian Human-Computer Communications Society, 1995.

[146] A. Rangarajan, H. Chui, and F. L. Bookstein. The softassign procrustes matching algorithm. In *Information Processing in Medical Imaging*, pages 29–42, London, UK, 1997. Springer-Verlag.

[147] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific Publishing Company, 1989.

[148] M. Rochery, I. H. Jermyn, and J. Zerubia. Phase field models and higher-order active contours. In *IEEE International Conference in Computer Vision*, Beijing, China, octobre 2005.

[149] K. Rohr, M. Fornefett, and H. S. Stiehl. Spline-based elastic image registration: integration of landmark errors and orientation attributes. *Computer Vision and Image Understanding*, 90(2):153–168, 2003.

[150] F. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.

[151] A. Rosenfeld and J. Pfaltz. Sequential operations in digital picture processing. *Journal of the Association for Computing Machinery*, 13(4):471–494, October 1966.

[152] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based space. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages II: 699–704, 2003.

[153] M. Rousson and N. Paragios. Shape Priors for Level Set Representations. In *European Conference on Computer Vision*, pages II:78–93, Copenhangen, Denmark, 2002.

[154] M. Rousson and N. Paragios. Prior knowledge, level set representations and visual grouping. *International Journal of Computer Vision*, (in press).

[155] M. Rousson, N. Paragios, and R. Deriche. Implicit active shape models for 3d segmentationin mri imaging. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2004.

[156] E. Rouy and A. Tourin. A Viscosity Solutions Approach to Shape-from-Shading. *SIAM Journal on Numerical Analysis*, 29:867–884, 1992.

[157] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18:712–721, 1999.

[158] S. Sain and D. Scott. On locally adaptive density estimation. *J. Am. Statist. Assoc.*, 91:1525–1534, 1996.

[159] S. R. Sain. Multivariate locally adaptive density estimation. *Comput. Stat. Data Anal.*, 39(2):165–186, 2002.

[160] B. Schölkopf and A. J. Smola. A short introduction to learning with kernels. *Proceedings of the Machine Learning Summer School*, pages 41–64, 2003.

[161] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, new york, chichester edition, 1992.

[162] T. Sederberg and S. Parry. Free-form deformation of solid geometric models. *Proceedings SIGGRAPH '86*, 20:151–160, 1986.

[163] J. Sethian. A Review of the Theory, Algorithms, and Applications of Level Set Methods for Propagating Interfaces. *Cambridge University Press*, pages 487–499, 1995.

[164] J. Sethian. *Level Set Methods*. Cambridge University Press, 1996.

[165] J. Sethian. Fast Marching Methods. *SIAM Review*, 41:199–235, 1999.

[166] E. Simoncelli. Bayesian multi-scale differential optical flow. In B. Jähne, H. Haussecker, and P. Geissler, editors, *Handbook of Computer Vision and Applications*, volume 2, chapter 14, pages 397–422. Academic Press, San Diego, April 1999.

[167] S. Soatto and A. J. Yezzi. DEFORMOTION: Deforming motion, shape average and the joint registration and segmentation of images. In *ECCV (3)*, pages 32–57, 2002.

[168] M. Sofka, G. Yang, and C. V. Stewart. Simultaneous covariance driven correspondence (CDC) and transformation estimation in the expectation maximization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 18–23 June 2007.

[169] J. Solem. *Variational Problems and Level Set Methods in Computer Vision - Theory and Applications*. PhD thesis, Centrum Scientiarum Mathematicarum, Lund University., 2006.

[170] M. Šonka, V. Hlaváč, and R. D. Boyle. *Image Processing, Analysis and Machine Vision*. Chapman and Hall, London, UK, first edition, 1993.

[171] L. Staib and S. Duncan. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:1061–1075, 1992.

[172] L. H. Staib and J. S. Duncan. Model-based deformable surface finding for medical images. *IEEE Trans. Medical Imaging*, 15(5):720–731, 1996.

[173] C. Stewart. Covariance-based registration. Technical report, Department of Computer Science, Rensselaer Polytechnic Institute, June 2002.

[174] C. Stewart. *The Handbook of Mathematical Models in Computer Vision*, chapter Uncertainty-Driven Point-Based Image Registration. Springer, 2005.

[175] C. Stewart, C.-L. Tsai, and B. Roysam. The dual bootstrap ICP algorithm with application to retinal image registration. *IEEE Trans. Med. Img.*, 22:1379–1394, 2003.

[176] C. V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, 1999.

[177] G. Sundaramoorthi, A. Yezzi, and A. Mennucci. Sobolev active contours. *IEEE International Conference in Computer Vision*, 73(3):345–366, July 2007.

[178] G. Szekely, A. Kelemen, C. Brechbuhler, and G. Gerig. Segmentation of 3d objects from mri volume data using constrained elastic deformations of flexible fourier surface models. In *Computer Vision, Virtual Reality and Robitics in Medicine, CVRMed95*, pages XX–YY, 1995.

[179] R. Szeliski, D. Tonnesen, and D. Terzopoulos. Modeling surfaces of arbitrary topology with dynamic particles. In IEEE, editor, *Computer Vision and Pattern Recognition (CVPR'93)*, page 82–87, 1993.

[180] J.-P. Tarel and D. B. Cooper. A new complex basis for implicit polynomial curves and its simple exploitation for pose estimation and invariant recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 111–117, Santa Barbara, California, USA, June 1998.

[181] M. Taron, C. Ghys, and N. Paragios. Uncertainties-driven surface morphing: The case of photo-realistic transitions between facial expressions. In *ICPR*, pages 889–892, 2006.

[182] M. Taron, N. Paragios, and M.-P. Jolly. Modelling shapes with uncertainties: Higher order polynomials, variable bandwidth kernels and non parametric density estimation. In *IEEE International Conference in Computer Vision*, pages 1659–1666, 2005.

[183] M. Taron, N. Paragios, and M.-P. Jolly. Uncertainty-driven non-parametric knowledge-based segmentation: The corpus callosum case. In *IEEE Workshop in Variational and Level Set Methods*, pages 198–207, 2005.

[184] M. Taron, N. Paragios, and M.-P. Jolly. From uncertainties to statistical model building and segmentation of the left ventricle. In *IEEE Mathematical Methods in Biomedical Image Analysis*, October 2007. In conjunction with ICCV 2007.

[185] M. Taron, N. Paragios, and M.-P. Jolly. Registration with uncertainties and statistical modeling of shapes with variable metric kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to appear.

[186] T. Tasdizen, J. Tarel, and D. Cooper. Improving the stability of algebraic curves for applications. *IEEE Transactions on Image Processing*, 9(3):405–416, Mar. 2000.

[187] G. Taubin, F. Cukierman, S. Sullivan, J. Ponce, and D. J. Kriegman. Parameterized families of polynomials for bounded algebraic curve and surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):287–303, 1994.

[188] G. Terrell and D. Scott. Variable kernel density estimation. *Ann. Statist.*, 20:1236–1265, 1992.

[189] D. Terzopoulos. On Matching Deformable Models to Images. Technical Report 60, Schlumberger Palo Alto Research, 1986.

[190] D. Terzopoulos and K. Fleischer. Deformable Models. *The Visual Computer*, 4:306–331, 1988.

[191] D. Terzopoulos and D. Metaxas. Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):703–714, 1991.

[192] P. Thévenaz and M. Unser. Precision isosurface rendering of 3-D image data. *IEEE Transactions on Image Processing*, 12(7):764–775, July 2003.

[193] D. W. Thompson. *On Growth and Form: The Complete Revised Edition*. Dover, 1917.

[194] A. Tikhonov. *Ill-Posed Problems in Natural Sciences*. Coronet, 1992.

[195] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society*, 61(3):611–622, 1999.

[196] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, A. Grimson, and A. Willsky. Model-based Curve Evolution Technique for Image Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 463–468, 2001.

[197] G. Turk and J. ÒBrien. Shape Transformation Using Variational Implicit Functions. In *ACM SIGGRAPH*, pages 335–342, 1999.

[198] M. Üzümcü, A. Frangi, J. Reiber, and B. Lelieveldt. Independent component analysis in statistical shape models. In *SPIE Medical Imaging*, volume 5032, pages 375–383, 2003.

[199] M. Vasilescu and D. Terzopoulos. Adaptive meshes and shells: Irregular triangulation, discontinuities, and hierarchical subdivision. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR*, pages 829–832, Los Alamitos, California, 15–18 1992. IEEE Computer Society.

[200] B. C. Vemuri, S. M. Rahman, and J. Li. Multiresolution adaptive k-means algorithm for segmentation of brain mri. In *ICSC '95: Proceedings of the Third International Computer Science Conference on Image Analysis Applications and Computer Graphics*, pages 347–354, London, UK, 1995. Springer-Verlag.

[201] B. J. H. Verwer. Local distances for distance transformations in two and three dimensions. *Pattern Recogn. Lett.*, 12(11):671–682, 1991.

[202] L. Vese and T. Chan. A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. *International Journal of Computer Vision*, 50:271–293, 2002.

[203] P. Viola and W. Wells. Aligment by Maximization of Mutual Information. In *IEEE International Conference in Computer Vision*, pages 16–23, 1995.

[204] C. Walder, O. Chapelle, and B. Schölkopf. Implicit surface modelling as an eigenvalue problem. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 936–939, New York, NY, USA, 2005. ACM Press.

[205] M. Wand and M. Jones. *Kernel Smoothing*. Chapman & Hall, 1995.

[206] W. Welch and A. Witkin. Free-form shape design using triangulated surfaces. In *ACM SIGGRAPH*, pages 247–256, 1994.

[207] W. Wells, R. Kikinis, W. Grimson, and F. Jolesz. Adaptive segmentation of mri data. *IEEE Transactions on Medical Imaging*, 15:429–442, 1996.

[208] D. Werman and D. Weinshall. Similarity and affine invariant distances between 2d points sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):810–814, 1995.

[209] Z. Xie and G. Farin. Deformation with hierarchical b-splines. In *Mathematical Methods for Curves and Surfaces*, pages 545–554. Vanderbilt University, 2001.

[210] Z. Xie and G. E. Farin. Image registration using hierarchical b-splines. *IEEE Trans. Vis. Comput. Graph.*, 10(1):85–94, 2004.

[211] L. Yatziv, A. Bartesaghi, and G. Sapiro. O(N) implementation of the fast marching algorithm. *Journal of Computational Physics*, 212:393–399, Mar. 2006.

[212] A. Yezzi, A. Tsai, and A. Willsky. A Statistical Approach to Snakes for Bimodal and Trimodal Imagery. In *IEEE International Conference in Computer Vision*, pages 898–903, 1999.

[213] H. Yoshida, B. Keserci, and K. Doi. Computer-aided diagnosis of pulmonary nodules in chest radiographs: A wavelet-based snake approach. In *CBMS*, pages 258–263, 1998.

[214] A. Yuille, P. Hallinan, and D. Cohen. Feature extration from faces using deformable templates. *International Journal of Computer Vision*, 8:99–111, 1992.

[215] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.

[216] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging.*, 20:45–57, 2001.

[217] Z. Zhang. Iterative Point Matching for Registration of Free-Form Curves and Surfaces. *International Journal of Computer Vision*, 13:119–152, 1994.

[218] S. Zhu, T. Lee, and A. Yuille. Region Competition: Unifying Snakes, Region Growing Energy/Bayes/MDL for Multi-band Image Segmentation. In *IEEE International Conference in Computer Vision*, pages 416–423, 1995.

[219] S. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884–900, 1996.

[220] W. Zhu and T. Chan. Stability for shape comparison model. Technical report, UCLA-CAM, August 2003.

Publication of the author

Journal:

- **Registration with Uncertainties and Statistical Modeling of Shapes with Variable Metric Kernels.** M. Taron, N. Paragios & M.-P. Jolly.
  IEEE Transactions on Pattern Analysis and Machine Intelligence, (to appear).

Conferences:

- **From Uncertainties to Statistical Model Building and Segmentation of the Left Ventricle.**
  M. Taron, N. Paragios & M.-P. Jolly.
  In Mathematical Methods in Biomedical Image Analysis (MMBIA). In conjunction with ICCV, 2007.

- **Uncertainties-driven Surface Morphing: The case of Photo-realistic Transitions between Facial Expressions.** M. Taron, C. Ghys & N. Paragios.
  In 18th International Conference on Pattern Recognition (ICPR), Hong Kong, 2006.

- **Uncertainty-driven Non-parametric Knowledge-based Segmentation: The Corpus Callosum Case.**
  M. Taron, N. Paragios & M.-P. Jolly.
  In 3rd ICCV workshop on Variational Geometric and Level Set Methods (VLSM) 2005.

- **Modelling Shapes with Uncertainties : Higher Order Polynomials, Variable Bandwidth Kernels and Non-Parametric Density Estimation.** M. Taron, N. Paragios & M.-P. Jolly.

In IEEE International Conference in Computer Vision (ICCV), 2005.

- **Active Shape Models & Segmentation of the Left Ventricle in Echocardiography.**
  N. Paragios, M.-P. Jolly, M. Taron & R. Ramaraj.
  In International Conference on Scale Space Theories and PDEs methods in Computer Vision, 2005.

Book Chapter:

- **On the Represantation of Shapes Using Implicit Functions.**
  N. Paragios, M. Taron, X. Huang, M. Rousson & D. Metaxas.
  In Statistics and Analysis of Shapes (Krim & Yezzi), Springer Verlag, 2006.

Résumé

Nous avons pu observer récemment d'importants progrès dans les techniques d'imageries médicales qui ont été accompagnés par le développement d'outils informatiques de prévisualisation et d'aide automatique au diagnostic. La réalisation de tels outils nécessite généralement la création d'un modèle mathématique capable de représenter les organes et dont la construction est divisée en trois étapes : (i) choisir et extraire les structures à étudier, (ii) choisir un modèle mathématique adapté à la représentation de ces structures particulières, (iii) estimer les variations des paramètres du modèle ainsi choisi. Cette thèse aborde donc chacune de ces taches de façon originale. La modélisation des organes est décrite au travers de déformations et nécessite une étape préalable de recalage de forme. Ceci implique la définition d'une forme de référence ainsi que d'un ensemble de déformations. Cette thèse introduit l'utilisation d'incertitudes sur le recalage de formes : définies a l'aide d'une matrice de covariance dans l'espace des déformations, elles indiquent localement la fiabilité du recalage obtenu. Ensuite vient la modélisation des variations de formes, obtenue à partir d'un ensemble d'apprentissage représentant différentes instances de l'organe étudié. Cette thèse apporte à la phase de modélisation des déformations, les informations sur les erreurs de recalage au travers de la propagation des incertitudes. La contribution finale de la thèse touche à la segmentation de ces structures par un modèle déformable, guidé par le modèle de forme sur lequel les incertitudes dues au modèle sont évaluées. La segmentation cardiaque du ventricule gauche en imagerie scanner, ainsi que le corps calleux en imagerie à résonance magnétique ont été considérés pour démontrer les performances de cette approche.

Abstract

In the recent years, we have witnessed a revolution on new non-invasive means for human and biological tissues imaging. The use of computer aided-techniques has emerged naturally as an efficient pre-screening and post treatment evaluation procedure. This often involves mathematical modeling of organs which usually refers to the three following steps: (i) determine/extract the structure of interest, (ii) provide a mathematical model to describe these structures and (iii) estimate the variations of the parameters in the proposed model. In this thesis we propose novel means to address the above tasks. The modeling of organs is performed with the description of the deformations. It therefore involves shape registration with the definition of a reference shape and a deformations space. In this thesis we introduce the use of uncertainties in the registration, defined as covariance matrices in the deformations space which indicate the amount of confidence in the obtained registration. The next step consists in modeling the shape variations based on a training set representing various instances of the organ under study. In this thesis we extend the state of the art that does not account for registration errors and introduce a method that propagates registration uncertainty to the modeling step. The last contribution of the thesis is in the area of knowledge based segmentation and consists of introducing a segmentation-by-deformation approach where the use of uncertainties both in the model as well as image space are considered. Segmentation of the cardiac left ventricle on CT scan and of the corpus callosum on MR-images using the above mentioned-methods are considered as applications to demonstrate the extreme potentials of our approach.